OM protein - nucleic search, using frame plus_p2n model

            Copyright   GenCore version 5.1.6
            Copyright (c) 1993 - 2003 Compugen Ltd.

Run on: December 17, 2003, 18:06:54 ; Search time 3970 Seconds
                         (without alignments)
                         2947.142 Million cell updates/sec

Title: US-10-054-313-1
Perfect score: 1546
Sequence: 1 MSWLLFLAHRVALAALPCRR.........FIGNEBADRLAREGAKQSED 286

Scoring table: BLOSUM62
            Xgapop 10.0 , Xgapext 0.5
            Ygapop 10.0 , Ygapext 0.5
            Fgapop 6.0 , Fgapext 7.0
            Delop   6.0 , Delext   7.0

Searched: 2888711 segs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US1005s313/runat_17122003_150743_24179/app_query.fasta_1.455
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10054313 @CGN_1_1_3508 @runat_17122003_150743_24179 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -THREADS=30 -TIMEOUT=30 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :   GenEmbl:*
        1: gb_ba:*
        2: gb_htg:*
        3: gb_in:*
        4: gb_om:*
        5: gb_ov:*
        6: gb_pat:*
        7: gb_ph:*
        8: gb_pl:*
        9: gb_pr:*
       10: gb_ro:*
       11: gb_sts:*
       12: gb_sy:*
       13: gb_un:*
       14: gb_vi:*
       15: em_ba:*
       16: em_fun:*
       17: em_hum:*
       18: em_in:*
       19: em_mu:*
       20: em_om:*
       21: em_or:*
       22: em_ov:*
       23: em_pat:*
       24: em_ph:*
       25: em_pl:*
       26: em_ro:*
       27: em_sts:*
       28: em_un:*
       29: em_vi:*
       30: em_htg_hum:*
       31: em_htg_inv:*
       32: em_htg_other:*
       33: em_htg_mus:*
       34: em_htg_pln:*
       35: em_htg_rod:*
       36: em_htg_mam:*
       37: em_htg_vrt:*
       38: em_sy:*
       39: em_htgo_hum:*
       40: em_htgo_mus:*
       41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                                SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1526	98.7	1158	9	BC002973	BC002973 Homo sapi
2	1526	98.7	1168	9	AF039652	AF039652 Homo sapi
3	1526	98.7	1604	9	AK075490	AK075490 Homo sapi
4	1522	98.4	1113	9	AF048994	AF048994 Homo sapi
5	1522	98.4	1147	9	AF048995	AF048995 Homo sapi
6	1518	98.2	861	9	HSAJ4117	AJ224117 Homo sapi
7	1457.5	94.3	2120	9	AK096913	AK096913 Homo sapi
8	1197.5	77.5	1430	10	AF048993	AF048993 Mus muscu
9	1193.5	77.2	1409	10	BC019411	BC019411 Mus muscu
10	1174	75.9	178965	9	AC098850	AC098850 Homo sapi
11	1174	75.9	192337	9	AC022596	AC022596 Homo sapi
12	1165	75.4	211403	9	AC126352	AC126352 Homo sapi
13	1150.5	74.4	207418	9	AC090774	AC090774 Homo sapi
14	1145.5	74.1	176040	9	AC107926	AC107926 Homo sapi
15	1016	65.7	131239	9	HS667H12	AL035414 Human DNA
16	956	61.8	2024	9	AY063502	AY063502 Cercopith
17	923	59.7	1992	9	AY063503	AY063503 Macaca as
18	882	57.1	1293	5	CHKESTFL25	D26340 Gallus gal
19	816	52.8	2165	9	AF448495	AF448495 Callithri
20	755.5	48.9	1911	9	AK057473	AK057473 Homo sapi
21	755.5	48.9	3471	9	BC041488	BC041488 Homo sapi
22	423.5	27.4	1153	3	AY089374	AY089374 Drosophil
23	375	24.3	1460	2	AF032921	AF032921 Drosophil
24	373.5	24.2	1117	3	TBU74470	U74470 Trypanoso
25	373.5	24.2	141775	3	AC105378	AC105378 Trypanoso
26	356.5	23.1	37983	2	AC020394	AC020394 Drosophil
27	356.5	23.1	85651	2	AC005463	AC005463 Drosophil
28	356.5	23.1	86889	3	AC005448	AC005448 Drosophil
29	356.5	23.1	139410	3	AC007084	AC007084 Drosophil
30	356.5	23.1	254023	3	AE003839	AE003839 Drosophil
31	331.5	21.4	34828	2	U41994	U41994 Caenorhabdi
32	312	20.2	16143	2	AL356960	AL356960 Homo sapi
33	303.5	19.6	2277	2	TRFRNH1A	L18916 Crithidia f
34	300.5	19.4	56083	2	AC108488	AC108488 Homo sapi
35	276	17.9	39512	8	SPBC336	SPBC336 S.pombe c
36	272	17.6	1557	8	AF048992	AF048992 Schizosac
37	272	17.6	119211	6	AX067465	AX067465 Sequence
38	268	17.3	231	6	BD118436	BD118436 EST and e
39	261.5	16.9	207104	2	AC136986	AC136986 Mus muscu
40	258	16.7	107893	5	BX276180	BX276180 Zebrafish
41	255.5	16.5	231062	2	AC109110	AC109110 Rattus no
42	255.5	16.5	246539	2	AC125638	AC125638 Rattus no
43	254	16.4	683	2	AF542055	AF542055 Leishmani
44	254	16.4	40397	8	SC9959	Z49939 S.cerevisia
45	253	16.4	300600	1	AP005369	AP005369 Thermosyn

                                ALIGNMENTS

RESULT 1

```
BC002973
LOCUS       BC002973                1158 bp    mRNA    linear   PRI 12-JUL-2001
DEFINITION  Homo sapiens, ribonuclease H1, clone MGC:2019 IMAGE:3537074, mRNA,
            complete cds.
ACCESSION   BC002973
VERSION     BC002973.1  GI:12804228
KEYWORDS    MGC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1158)
AUTHORS     Strausberg,R.
TITLE       Direct Submission
JOURNAL     Submitted (06-FEB-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
REMARK      NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Genome Sequence Centre,
            BC Cancer Agency, Vancouver, BC, Canada
            info@bcgsc.bc.ca
            Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
            Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
            Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
            Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
            Ness, Pawan Pandoh, Anna-liisa Prabhu, Parvaneh Saeedi, Jacqueline
            Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott,
            Michael Thorne, Miranda Tsai, Natasja van den Bosch, Jill Vardy,
            George Yang, Scott Zuyderduyn, Marco Marra.
            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 8 Row: 1 Column: 11
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 3004980.
FEATURES             Location/Qualifiers
     source          1..1158
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="LocusID:6040"
                     /db_xref="taxon:9606"
                     /clone="MGC:2019 IMAGE:3537074"
                     /tissue_type="Lung, small cell carcinoma"
                     /clone_lib="NIH_MGC_7"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
     CDS             71..931
                     /codon_start=1
                     /product="ribonuclease H1"
                     /protein_id="AAH02973.1"
                     /db_xref="GI:12804229"
                     /translation="MSWLLFLAHRVALAALPCRRGSRGPGMFYAVRRGRKTGVFLTWN
                     ECRAQVDRFPAARFKKFATEDEAWAFVRKSASPEVSEGHENGHGQESEAKASKRLREP
                     LDGDGHESAEPYAKHMKPSVEPAPPVSRDTFSYMGDFVVVYTDGCCSSNGRRRPRAGI
                     GVYWGPGHPLNVGIRLPGRQTNQRAEIHAACKAIEQAKTQNINKLVLYTDSMFTINGI
                     TNWVQGWKKGWKTSAGKEVINKEDFVALERLTQGMDIQWMHVPGHSGFIGNEEADRL
                     AREGAKQSED"
BASE COUNT      317 a    248 c    349 g    244 t
ORIGIN
```

```
Alignment Scores:
Pred. No.:                          1158
Score:              1.61e-101       Matches:          282
Percent Similarity:     1526.00     Conservative:       2
Best Local Similarity:  99.30%      Mismatches:         2
Query Match:            98.60%      Indels:             0
DB:                     98.71%      Gaps:               0
                            9
```

```
US-10-054-313-1 (1-286) x BC002973 (1-1158)

Qy       1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaAlaLeuProCysArgArg 20
Db      71 ATGAGCTGGCTTCTGTTCCTGGCCCACAGAGTCGCTCTGGCCGCTCTGCCTTGCCTGCGGCGC 130

Qy      21 GlySerArgGlyPheGlyMetPheTyrAlaValArgArgGlyArgLysThrGlyValPhe 40
Db     131 GGCTCTCGCGGCGGGTTCGGGATGTTCTATGCCGTGAGGAGGGGCCCGCAAGACCGGGGTCTT 190

Qy      41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys 60
Db     191 CTGACCTGGAATGAGTGCAGAGCACAGGTGGACCGGTTCCCTGCTGCCAGATTTAAGAAG 250

Qy      61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80
Db     251 TTTGCCACAGAGGATGAGGCCTGGGCCTTTGTCAGGAAATCTGCAAGTCCTGAAGTTTCA 310

Qy      81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100
Db     311 GAAGGGCATGAAAATCAACATGGACAAGAATCGGAGGGCCAAGAGCCAGCAAGCGACTCCGT 370

Qy     101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120
Db     371 GAGCCACTGGATGGAGATGGACATGAAAGCGCAGAGCCGTATGCAAAGCACATGAAGCCG 430

Qy     121 SerValGluProAlaProProValSerArgAspThrPheSerTyrMetGlyAspPheVal 140
Db     431 AGCGTGGAGCCGGCGCCTCCAGTTAGCAGAGACACGTTTTCCTACATGGGAGACTTCGTC 490

Qy     141 ValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgGlyArgArgGlyAlaGlyIle 160
Db     491 GTCGTCTACACTGATGGCTGCTGCTCCAGTAATGGCCGTAGAAGGCCGCGAGCAGGAAATC 550

Qy     161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGln 180
Db     551 GGCGTTTACTGGGGGCCAGGCCATCCTTAAATGTAGGCATTAGACTTCCTGGGCGGCAG 610

Qy     181 ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaIleGluGlnAlaLysThrGln 200
Db     611 ACAAACCAAAGAGCGGAAATTCATGCAGCCTGCAAAGCCATTGAACAAGCAAAGACTCAA 670

Qy     201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220
Db     671 AACATCAATAAACTGGTTCTGTATACAGATAGCATGTTTACGATAAATGGTATAACTAAC 730

Qy     221 TrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGluValIleAsn 240
Db     731 TGGGTTCAAGGTTGGAAGAAAATGGGTGGAAGACAAGTGCAGGGAAAGAGGTGATCAAC 790

Qy     241 LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHis 260
Db     791 AAAGAGGACTTTGTGGCACTGGAGAGCTTACCCAGGGGATGGACATTCAGTGGATGCAT 850

Qy     261 ValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgGluGly 280
Db     851 GTTCCTGGTCATTCGGGATTTATAGGCAATGAAGAGAGCTGACGAGATTAGCCAGAGAGGA 910

Qy     281 AlaLysGlnSerGluAsp 286
Db     911 GCTAAACAATCGGAAGAC 928
```

```
RESULT 2
AF039652
LOCUS       AF039652                1168 bp    mRNA    linear   PRI 02-APR-1998
DEFINITION  Homo sapiens ribonuclease H type II mRNA, complete cds.
ACCESSION   AF039652
VERSION     AF039652.1  GI:3004980
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
Db    502  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
           GTCGTCCTACACTGATGGCTGCTGCTCCAGTAATGGGCGTAGAAGGCCGCGAGCAGGAATC 561

Qy    161  GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGln 180
Db    562  GGCGTTTACTCGGGGCCAGGCCATCCTTAAATGTAGGCATTAGACTTCCTGGGCGGCCAG 621

Qy    181  ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaIleGluGlnAlaLysThrGln 200
Db    622  ACAAACCAAAGAGCGAAATTCATCAGCCTGCAAAGCCATTGAACAAGCAAAGACTCAA 681

Qy    201  AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220
Db    682  AACATCAATAAACTGGTTCTGTATACAGACAGTATGTTTACGATAAATGGTATAACTAAC 741

Qy    221  TrpValGlnGlyValAsnGlyTrpLysThrSerAlaGlyValLysGluValIleAsn 240
Db    742  TGGGTTCAAGGTTGGAAGAAAAATGGTGGAAGACAAGTGCAGGGAAAGAGGTGATCAAC 801

Qy    241  LysGluAspPheValAlaLeuArgLeuThrGlnGlyMetAspIleGlnTrpMetHis 260
Db    802  AAAGAGGACTTTGTGGCACTGGAGAGGCTTACCCAGGGGATGGACATTCAGTGGATGCAT 861

Qy    261  ValProGlyHisSerGlyPheIleGlyAsnGluAlaAspArgLeuAlaArgGluGly 280
Db    862  GTTCCTGGTCATTCGGGATTTATAGGCAATGAAGAAGCTGACAGATTAGCCAGAGAAGGA 921

Qy    281  AlaLysGlnSerGluAsp 286
Db    922  GCTAAACAATGGAAGAC 939
```

RESULT 3
AK075490
LOCUS   AK075490   1604 bp   mRNA   linear   PRI 03-SEP-2002
DEFINITION Homo sapiens cDNA PSEC0185 fis, clone OVARC1002091, highly similar
            to Homo sapiens ribonuclease H type II mRNA.
ACCESSION   AK075490
VERSION   AK075490.1 GI:22761667
KEYWORDS   oligo capping; fis (full insert sequence).
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
AUTHORS   Ota,T., Nishikawa,T., Suzuki,Y., Kawai-Hio,Y., Hayashi,K.,
            Ishii,S., Saito,K., Yamamoto,J., Wakamatsu,A., Nagai,T.,
            Nakamura,Y., Nagahari,K., Sugano,S. and Isogai,T.
TITLE   HRI human cDNA sequencing project
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 1604)
AUTHORS   Isogai,T. and Yamamoto,J.
TITLE   Direct Submission
JOURNAL   Submitted (20-MAR-2002) Takao Isogai, Helix Research Institute,
            Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT   HRI human cDNA sequencing project; cDNA 5'- & 3'-end one pass
            sequencing, clone selection and full insert sequencing: Helix
            Research Institute (supported by Japan Key Technology Center etc.);
            cDNA library construction: Institute of Medical Science, University
            of Tokyo, Laboratory of Genome Structure, Human Genome Center.
FEATURES   Location/Qualifiers
     source   1..1604
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="OVARC1002091"
                     /tissue_type="ovary, tumor tissue"
                     /clone_lib="OVARC1"
                     /note="cloning vector: pME18SFL3"
BASE COUNT   426 a   343 c   460 g   375 t
ORIGIN

Alignment Scores:

FEATURES   Location/Qualifiers
     source   1..1168
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17p11.2"
     CDS   82..942
                     /function="ribonuclease specific for the RNA strand of a
                     DNA-RNA hybrid"
                     /note="RNAase HII"
                     /codon_start=1
                     /product="ribonuclease H type II"
                     /protein_id="AAC09261.1"
                     /db_xref="GI:3004981"
                     /translation="MSWLLFLAHRVALAALPCRRGSRGFGMFYAVRRGRKTGVFLTWN
                     ECRAQVDRFPAARFKFATEDEAWAFVRKSASPEVSEGHENQHGQESEAKASKRLREP
                     LDGDHESAEPYAKHMKPSVEPAPPVSRDTFSYMGDFFVVYTDGCCSSNGRRRPRAGI
                     GVYWGPGHPLNVGIRLPGRQNQRABIHAACKAIEQAKTQNINKILVLYTDSMFTTNGI
                     TNWVQGWKKNGWKTSAGKEVINKEDFVALERLTQGMDIQWMHVPGHSGFIGNEEADRL
                     AREGAKQSED"

BASE COUNT   325 a   249 c   349 g   245 t
ORIGIN

Alignment Scores:
Pred. No.:   Length: 1168
Score: 1526.00   Matches: 282
Percent Similarity: 99.30%   Conservative: 2
Best Local Similarity: 98.60%   Mismatches: 2
Query Match: 98.71%   Indels: 0
DB: 9   Gaps: 0

US-10-054-313-1 (1-286) x AF039652 (1-1168)

```
Qy      1  MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaAlaLeuProCysArgArg 20
Db     82  ATGAGCTGGCTTCTGTTCCTGGCCCACAGAGTCGCCCTTGGCCGCCGCGGCCGCCGCGC 141

Qy     21  GlySerArgGlyPheGlyMetPheTyrAlaValArgArgGlyArgLysThrGlyValPhe 40
Db    142  GGCTCTCGCGGGTTCGGGATGTTCTATGCCGTGAGGAGGGGCCGCAAGACCGGGGTCTTT 201

Qy     41  LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys 60
Db    202  CTGACCTGGAATGAGTGCAGAGCACAGGTGGACCGGTTTCCTGCTGCCAGATTAAGAAG 261

Qy     61  PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80
Db    262  TTTGCCACAGAGGATGAGGCCTGGGCCTTTGTCAGGAAATCTGCAAGCCCGGAAGTTTCA 321

Qy     81  GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100
Db    322  GAAGGGCATGAAAATCAACATGGACAAGAATCGGAGGCGAAAGCCAGCAAGCCGACTCCGT 381

Qy    101  GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120
Db    382  GAGCCACTGGATGGAGATGGACATGGACAAGAGCCGTATGCAAAGCACATGAAGCCG 441

Qy    121  SerValGluProAlaProProValSerArgAspThrPheSerTyrMetGlyAspPheVal 140
Db    442  AGCGTGGAGCCGGCGCCTCCAGTTAGCAGAGACACGTTTTCCTACATGGGAGACTTCGTC 501

Qy    141  ValValThrAspGlyCysCysSerSerAsnGlyArgArgArgProArgAlaGlyIle 160
```

```
Pred. No.:              2.31e-101      Length:           1604
Score:                  1526.00        Matches:          282
Percent Similarity:     99.30%         Conservative:     2
Best Local Similarity:  98.60%         Mismatches:       2
Query Match:            98.71%         Indels:           0
DB:                                    Gaps:             0

US-10-054-313-1 (1-286) x AK075490 (1-1604)

Qy      1  MetSerTrpLeuPheLeuAlaHisArgValAlaAlaLeuAlaAlaLeuProCysArgArg 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61  ATGAGCTGGCTTCTCGTTCCTGCCCACAGAGTCGCCTTGCCCGCCCTTGCCCTGCCGCCGC 120

Qy     21  GlySerArgGlyMetPheTyrAlaValArgArgGlyTyrArgLysThrGlyValPhe 40
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121  GGCTCTCGCGGGTTCGGGATGTTCTATGCCGTGAGGAGGGGCCGCAAGACCGGGGTCTTT 180

Qy     41  LeuThrTrpAsnGluCysArgAlaGlnValAspArgProAlaAlaArgPheLysLys 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181  CTGACCTGGAATGAGTGCAGAGCACAGGTGGACCGGTTCCTGCTGCCAGATTTAAGAAG 240

Qy     61  PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241  TTTGCCACAGAGGATGAGGCCCTGGGCCTTGTCAGGAAATCTGCAAGCCCCGAAGTTTCA 300

Qy     81  GluGlyHisGluAsnGlnHisGlyTyrGlnEluGluValGlyAlaLysProGlyLysArgLeuArg 100
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301  GAAGGGCATGAAAATCAACATGGACAAGAAATCGGAGGCGAAAGCCAGCAAGCCGACTCCGT 360

Qy    101  GluProLeuAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120
              ||||||||||||||||||||||||| :::|||||||||||||||||||||||||
Db    361  GAGCCCCTGGATGGAGATGGACAAGAGCCAGCCGTATGCAAAGCACATGAAGCCG 420

Qy    121  SerValGluProProAlaProValSerArgAspThrPheSerTyrMetGlyAspPheVal 140
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421  AGCGTGGAGCCGCCGGCCCTCCAGTTAGCAGAGACACGTTTCCTACATGGGGAACTTCGTC 480

Qy    141  ValValTyrThrAspProVVVTCysSerSerAsnGlyArgArgProAlaGlyAlaGlyIle 160
              |||||||||||||||||||||::|::||||||||||||||||||||||||||||||||
Db    481  GTCGTCTACACTGATGGCTGTGCTCCAGTAATGGGCGTAGAAGGCCGCTAGAAGGCAGGAATC 540

Qy    161  GlyValTyrThrGlyProGlyHisProLeuAsnValGlyLyrIleArgGlnProGlyArgGln 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541  GGCGTTTACTGGGGCCCGGGCCATCCTTAAATGTAGGCATCATTAGACTTCCTGGCGGCAG 600

Qy    181  ThrAsnGlnInArgAlaGluIleHisAlaAlaCysLysAlaIleGluGlnHisAlaLysThrGln 200
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601  ACAAACCAAAGAGCGGGGAAATTCATGCAGCCTGCAAAGCCCATTGAACAAGCAAAGACTCAA 660

Qy    201  AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661  AACATCAATAAACTGGTTCTGTATACAGACACAGTGTTCACCATAAATGGTATAACTAAC 720

Qy    221  TrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysValIleAsn 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721  TGGGTTCAAGGTTGGAAGAAAAATGGGTGGAAGACAAGTGCAGGGAAAGAGGTGATCAAC 780

Qy    241  LysGluGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHis 260
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781  AAAGAGGAGGACTTTGTGCACTGGAGGGTTACCCAGGGGATGGACATTCAGTGGATGCAT 840

Qy    261  ValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgGluGlyLy 280
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841  GTTCCTGGGTCATTCGGGATTTATAGGCAATGAAGAAGCTGACAGATTAGCCAGAGAGGA 900

Qy    281  AlaLysGlnSerGluAsp 286
              ||||||||||||||
Db    901  GCTAAACAATCGGAAGAC 918
```

RESULT 4
AF048994
LOCUS   AF048994   1113 bp   mRNA   linear   PRI 21-NOV-1998
DEFINITION Homo sapiens ATCC158373 ribonuclease H1 (rnh1) mRNA, complete cds.

ACCESSION   AF048994
VERSION   AF048994.1   GI:2935441
KEYWORDS   .
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1113)
  AUTHORS   Cerritelli,S.M., Fedoroff,O.Y., Reid,B.R. and Crouch,R.J.
  TITLE   A common forty amino acid motif in eukaryotic RNases H1 and
            caulimovirus ORF VI proteins binds to duplex RNAs
  JOURNAL   Nucleic Acids Res. (1998) In press
REFERENCE   2 (bases 1 to 1113)
  AUTHORS   Cerritelli,S.M. and Crouch,R.J.
  TITLE   Cloning, expression, and mapping of ribonucleases H of human and
            mouse related to bacterial RNase HI
  JOURNAL   Genomics 53 (3), 300-307 (1998)
  MEDLINE   99017966
  PUBMED   9799596
REFERENCE   3 (bases 1 to 1113)
  AUTHORS   Cerritelli,S.M. and Crouch,R.J.
  TITLE   Direct Submission
  JOURNAL   Submitted (17-FEB-1998) Laboratory of Molecular Genetics, NICHD
            NIH, 6 Center Drive Building 6B Room 2B-231, Bethesda, MD 20892,
            USA
FEATURES   Location/Qualifiers
     source   1..1113
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /strain="ATCC158373"
                     /db_xref="ATCC:158373"
                     /db_xref="dbEST:AA356084"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17p11.2"
                     /cell_type="Jurkat T-cells VI"
     gene   1..1113
                     /gene="rnh1"
     CDS   28..888
                     /gene="rnh1"
                     /note="similar to bacterial RNases HI"
                     /codon_start=1
                     /product="ribonuclease H1"
                     /protein_id="AAC78563.1"
                     /db_xref="GI:2935442"
                     /translation="MSWFLFLAHRVALALAALPCRRGSRGFGMFYAVRRGRKTGVFLTWN
                     ECRAQVDRPPAARFKFATEDEAWAFVRKSASPEVSEGHENQHQESEAKASKRLREP
                     LDGDGHESAEPYAKHMKPSVEPAPPVSRDTFSYMGDFVVVYTDGCCSSNGRRRPRAGI
                     GVYWGPHGLPNVGIRLPGRQTNQRAEIHAACKAIEQAXTQNINKLVLYTDSMFTINGI
                     TNWVQGWKKNGWKTSAGKEVINKEDFVALERLTQGMDIQWMHVPGHSGFIGNEEADRL
                     AREGAKQSED"

BASE COUNT   311 a   232 c   333 g   237 t
ORIGIN

Alignment Scores:
Pred. No.:   2.99e-101   Length: 1113
Score: 1522.00   Matches: 281
Percent Similarity: 98.95%   Conservative: 2
Best Local Similarity: 98.25%   Mismatches: 3
Query Match: 98.45%   Indels: 0
DB: Gaps: 0

US-10-054-313-1 (1-286) x AF048994 (1-1113)

Qy   1 MetSerTrpLeuPheLeuAlaHisArgValAlaAlaLeuAlaAlaLeuProCysArgArg 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   28 ATGAGCTGGTTTCTGTTCCTGCCCACAGAGTCGCCTTGCCCGCCCTTGCCCTGCCGCCGC 87

Qy   21 GlySerArgGlyPheMetPheTyrAlaValArgArgGlyTyrArgLysThrGlyValPhe 40
              ||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db   88 GGCTCTCGCGGGTTCGGGATGTTCTATGCCGTGAGGAGGGGCCGCAAGACCGGGGTCTTT 147

Qy   41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgProAlaAlaArgPheLysLys 60
```

```
PUBMED       9799596
REFERENCE    3  (bases 1 to 1147)
AUTHORS      Cerritelli,S.M. and Crouch,R.J.
TITLE        Direct Submission
JOURNAL      Submitted (17-FEB-1998) Laboratory of Molecular Genetics, NICHD
             NIH, 6 Center Drive Building 6B Room 2B-231, Bethesda, MD 20892,
             USA

FEATURES             Location/Qualifiers
     source          1..1147
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /strain="ATCC159806"
                     /db_xref="ATCC:159806"
                     /db_xref="dbEST:AA312554"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17p11.2"
                     /cell_type="Jurkat T-cells VI"
     gene            1..1147
                     /gene="rnh1"
     CDS             62..922
                     /gene="rnh1"
                     /note="similar to bacterial RNases HI"
                     /codon_start=1
                     /product="ribonuclease H1"
                     /protein_id="AAC78564.1"
                     /db_xref="GI:2935444"
                     /translation="MSWFLFLAHRVALAALPCRRGSRGFMFYAVRRGRKTGVFLTWN
                     ECRAQVDRFPPAARFKKFATEDEAWAFVRKSASPEVSEGHENQKGQESEAKASKLREP
                     LDGDGHESAEPYAKHMKPSVEPAPPVSRDTFSYMGDFVVVYTDGCCSSNGRRRPRAGI
                     GVYWGPGHPLNVGIRLPGRQTNQRAEIHAACKAIEQAKTQNINKLVLYTDSMTINGI
                     TNWVQGWKKNGWKTSAGKEVINKEDFVALERLTQGMDIQWMHVPGHSGFIGNEEADRL
                     AREGAKQSED"

BASE COUNT     316 a    244 c    345 g    242 t
ORIGIN

Alignment Scores:
Pred. No.:            3.09e-101         Length:             1147
Score:               1522.00           Matches:             281
Percent Similarity:  98.95%            Conservative:          2
Best Local Similarity: 98.25%          Mismatches:            3
Query Match:         98.45%            Indels:                0
Db:                                    Gaps:                  0

US-10-054-313-1 (1-286) x AF048995 (1-1147)

Qy     1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaAlaLeuAlaAlaLeuProCysArgArg   20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    62 ATGAGCTGGTTTCTGTTCCTGGCCCACAGAGTCGCCGCCTTGGCCGCCTTGCCCTGCCGCCGC  121

Qy    21 GlySerArgGlyPheMetPheTyrAlaValArgArgGlyArgLysThrGlyValPhe   40
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   122 GGCTCTCGCGGGTTCATGTTCTACGCAGTGAGAAGAGGGCGCAAGACCGGGGTCTTT  181

Qy    41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys   60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   182 CTGACCTGGAATGAGTGCAGAGCACAGGTGGACCGGTTTCCTGCTGCCAGATTTAAGAAG  241

Qy    61 PheAlaAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer   80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   242 TTTGCCACAGAGGATGAGGCCTGGGCCTTTGTCAGGAAAATCTGCAAGCCCGGAAGTTTCA  301

Qy    81 GluGlyHisGluAsnGlnHisGlyGluAspLysGlnGlnSerGluAlaLysProGlyLysArgLeuArg  100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   302 GAAGGGCATGAAAATCAACATGACAAGAAATGGAGCGAAAGCCAAGCGACCTCGT  361

Qy   101 GluProLeuAspGlyAspGlyHisGluSerAlaGluProTyrAlaLysHisMetLysPro  120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   362 GAGCCACTGGATGGAGATGGACATGAAAGCCAGAGCCGATGCAAAGCACATGAAGCCG  421

Qy   121 SerValGluProAlaProProValSerArgAspThrPheSerTyrMetGlyAspPheVal  140
          ||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||||
Db   422 AGCGTGGAGCCCGCGCCTCCAGTTAGCAGAGACACGTTTTCCTACATGGGAGACTTCGTC  481
```

```
                    LDGDGHESABPYAKHMKPSVBPAPPVSRDTPSYMGDPVVVTDGCCSSNGRRRPRAGI
                    GVYWGPGHPLNVGIRLPGRQTNQRAEIHAACKAIEQAKTQNINKLVLYTDSMFTINGI
                    TNWVRGWKKNGWKTSAGKEVINKEDPVALERLTQGMDIQWMHVPGHSGFIGNEEADRL
                    AREGAKOSED"

BASE COUNT    236 a    189 c    263 g    173 t
ORIGIN

Alignment Scores:
Pred. No.:                          Length:          861
Score:                1518.00       Matches:         280
Percent Similarity:   99.30%        Conservative:      4
Best Local Similarity: 97.90%       Mismatches:        2
Query Match:          98.19%        Indels:            0
DB:                                 Gaps:              0

                                    4.37e-101

US-10-054-313-1 (1-286)  x  HSAJ4117 (1-861)

Qy    1  MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaAlaLeuProCysArgArg      20
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  ATGAGCTGGCTTCTGTTCCTGGCCCACAGAGTGGCCCTGGCCGCCCTGCCCTGCCGCCGC    60

Qy   21  GlySerArgGlyPheGlyMetPheTyrAlaValArgArgGlyArgLysThrGlyValPhe      40
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  GGCTCTCGCGGGTTCGGGATGTTTTATGCCGTGAGGAGGGGCCGCAAGACCGGGGTCTTT   120

Qy   41  LeuThrTrpAsnGluCysArgAlaGlnGlnAlaAspArgPheProAlaAlaArgPheLysLys   60
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  CTGACCTGGAATGAGTGCAGAGCACAAGTGGACCGGTTTCCTGCTGCCAGATTTAAGAAG   180

Qy   61  PheAlaThrAlaAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer      80
            ::|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181  TTTGCCACAGCAGATGAGGCCTGGGCCTTTGTCAGGAAATCTGCAAGCCCGGAAGTTTCA   240

Qy   81  GluGlyHisGluAsnGlnHisGlyArgGluSerGluAlaLysProGlyLysProLeuArg     100
            ||||||||||||||||||||::|||||::|||||||||||||||||||||||||||||||
Db  241  GAAGGGCATGAAAATCAACATGGACAGAGAATCGGAGGCAGGAAGCCAGCAAGCGACTCGT   300

Qy  101  GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro    120
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301  GAGCCACTGGATGGAGATGGACATGAAAGCCAGAGCCGTATGCAAAGACCATGAAGCCG    360

Qy  121  SerValGluProAlaProProValSerArgAspThrProSerTyrMetGlyAspPheVal    140
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361  AGCGTGGAGCCGGCGCCGCCTGTCAGTAGGCAGACACGTTTCCTACGGGGAGACTTCGTC   420

Qy  141  ValValTyrThrAspGlyCysSerSerAsnGlyArgArgArgLysProArgAlaGlyIle    160
            ||||||||||||||||||||||||::|||||||||||||||||||||||||||||||||
Db  421  GTCGTCTACACTGATGGCTGCTCGCTCCAGTAATGGGCGTAGAAGGCCGCGAGCAGGAATC   480

Qy  161  GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGln    180
            |||||||||||||||||||||||||||||||||||||||||||||||||||||::||||
Db  481  GGCGTTTACTGGGGCCCAGGCCATCCTTAAATGTGAGGCATTAGGACTTCCTGGGCGGCAG   540

Qy  181  ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaIleGluGlnAlaLysThrGln    200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541  ACAAACCAAAGAGCAGAAATTCATGCAGCCTGCAAAGCCATTGAACAAGCAAGACTCAA    600

Qy  201  AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn    220
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601  AACATCAATAAACTGGTTCTGTATACAGACAGTATGTTTACGATAAATGGTATAACTAAC   660

Qy  221  TrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGluValIleAsn    240
            |||||||||||||||||||::|||||||||||||||||||||||||||||||||||||||
Db  661  TGGGTTCGAGGTTGGAAGAAAAATGGGTGGAAGACAAGTCAGCAGGAAAGAGGTGATCAAC   720

Qy  241  LysGluAspPheValAlaLeuGluArgLeuThrGlnMetAspIleGlnTrpMetHis    260
            |||||||||||||||||||||||||||||||||||||||::|||||||||||||||||||
Db  721  AAAGAGGACTTTGTGCCACTGGAGCGGCTTACCCAGGGATGGACATTCAGTGGATGCAT   780

Qy  261  ValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgGluGly    280
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  781  GTTCCTGGTCATTCGGGATTTATAGGCAATGAAGAAGCTGACAGATTAGCCAGAGAAGGA   840

Qy  281  AlaLysGlnSerGluAsp                                          286
            |||||||||||||||||
Db  901  GCTAAACAATCCGGAAGAC                                        919
```

```
FEATURES          Location/Qualifiers
     source       1..861
                  /organism="Homo sapiens"
                  /mol_type="mRNA"
                  /db_xref="taxon:9606"
     gene         1..861
                  /gene="rnhII"
     CDS          1..861
                  /gene="rnhII"
                  /note="evolutionarily related to prokaryotic RNase HI and
                  RNase H domain of retroviral reverse transcriptase;
                  ribonuclease HII"
                  /codon_start=1
                  /product="RNase HII"
                  /protein_id="CAA11835.1"
                  /db_xref="GI:3043447"
                  /db_xref="SPTREMBL:O60857"
                  /translation="MSWLLFLAHRVALAAALPCRRGSRGFGMFYAVRRGRKTGVFLTWN
                  ECRAQVDRFPAARFKKFATEDEAWAFVRKSASPEVSEGHENQHGRESEAKASKRLREP
```

```
Qy   281 AlaLysGlnSerGluAsp 286
         ||||||||||||||||||
Db   841 GCTAAACAATCGGAAGAC 858


RESULT 7
AK096913
LOCUS       AK096913              2120 bp    mRNA    linear    PRI 15-JUL-2002
DEFINITION  Homo sapiens cDNA FLJ39594 fis, clone SKNSH2001875, highly similar
            to Homo sapiens ATCC158373 ribonuclease H1 (rnnh1) mRNA.
ACCESSION   AK096913
VERSION     AK096913.1  GI:21755512
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Kawakami,B., Sugiyama,A., Takemoto,M., Sugiyama,T., Irie,R.,
            Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y.,
            Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H.,
            Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K.,
            Wagatsuma,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A.,
            Oshima,A., Suzuki,Y., Sugano,S., Nagahari,K., Masuho,Y., Nagai,K.
            and Isogai,T.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2120)
  AUTHORS   Isogai,T. and Yamamoto,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team), 2-6-7
            Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
            (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
  COMMENT   NEDO human cDNA sequencing project supported by Ministry of
            Economy, Trade and Industry of Japan; cDNA full insert sequencing:
            Research Association for Biotechnology (RAB); cDNA library
            construction: Helix Research Institute (HRI) (supported by Japan
            Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
            HRI, and Biotechnology Center, National Institute of Technology and
            Evaluation; clone selection for full insert sequencing: HRI and
            RAB; annotation: HRI and RAB.
FEATURES             Location/Qualifiers
     source          1..2120
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="SKNSH2001875"
                     /cell_line="SK-N-SH"
                     /cell_type="neuroblastoma"
                     /clone_lib="SKNSH2"
                     /note="cloning vector: pME18SFL3"
BASE COUNT     546 a    476 c    624 g    474 t
ORIGIN


US-10-054-313-1 (1-286) x AK096913 (1-2120)

Alignment Scores:
Pred. No.:           1
Score:               2.77e-96          Length:            2120
Percent Similarity:  1457.50           Matches:           282
Best Local Similarity: 77.60%          Conservative:      2
Query Match:         77.05%            Mismatches:        2
DB:                  94.28%            Indels:            80
                     9                 Gaps:              1


Qy    1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaValAlaArgValAlaLeuAlaLeuProCysArgArg 20
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   30 ATGAGCTGGCTTCTTCGTTCCTGGCCCACAGAGTCGCCTTGGCCCGCCCTTGCCCTGCCGCCGCGCC 89

Qy   21 GlySerArgGlyPheGlyMetPheTyrAlaValArgArgGlyArgGlyThrGlyArgGlyValPhe 40
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   90 GGCTCCCGCGGGTTCGGGATGTTCTATGCCGTGAGGAGGGGCCGCAAGACCGGGGTCTTT 149

Qy   41 LeuThrTrp 43
        |||||||
```

```
Db   150 CTGACTTGGTGAGTGGCCGCGCCGCCGGCGGCGCCCTCCTGCGCGGCATCGGGGTCGAGGGA 209
                                                                                
Qy    43 ------------------------------------------------------------- 43

Db   210 GGCCGGAGCCCGGCTAACCCGGCGCGGCGGAGCCTGGCGGTCGCGGTCCGAGTCTGCGCCGCGC 269
                                                                                
Qy    43 ------------------------------------------------------------- 43

Db   270 TGTGGCTGCTGGTGGGTCGGGGGACCACCGGTCCTCGTTTTCCTGCGTCGGCCGGCGGGTTCCAGG 329
                                                                                
Qy    43 ------------------------------------------------------------- 43

Db   330 AGGCTCATGGCCTGGGACTGGAACGCACCGAAGACTTGTTTGAAATATCAGAAACGTCGC 389
                                                                                
Qy    44 ------AsnGluCysArgAlaGlnIleValAspArgArgPheProAlaAlaArgPheLysLysPh 61
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   390 GTAAAGGAATGAGTGCAGAGCAGGTGGACCGGTTCCTGCTGCCAGATTTAAGGAAGTT 449

Qy   61 eAlaThrGluAspGluAlaTrpAlaPheArgLysSerAlaSerProGluValSerGl 81
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   450 TGCCACAGAGGATGAGCCCTGGGCCTTTGTCAGGAAATCTGCAAGCCGGAAGTTTCAGA 509

Qy   81 uGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArgGl 101
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   510 AGGGCATGAAAATCAACATGGACACAGAATCGGAGGCGAAGCCCAGCAAGCACTCCGTGA 569

Qy  101 uProLeuAspGlyAlaAspGlyHisGluSerAlaGlnProTyrArgAlaLysHisMetLysProSe 121
        ||||||||||||||||||||||||||||||||::::||||||||||||||||||
Db   570 GCCACTGGATGGAGATGGACATGAAAGCCCAGAGCCGTATGCAAAGCACATGAAGCCGAG 629

Qy  121 rValGluProAlaProProValSerArgAspThrPheSerTyrMetGlyAspPheValVa 141
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   630 CGTGGAGCCGGCGCCCCTCCAGTTAGCACGAGACACGTTTCCTACATGGGAGACTTCGTCGT 689

Qy  141 lValTyrThrAspGlyCysSerSerAsnGlyArgArgLysProArgAlaGlyLyIleGl 161
        |||||||||||||||||||||||||||||||::::|||||||||||||||||||
Db   690 CGTCTACACTGATGGCTGCTCCAGTAATGGCCGTAGAGGCCCGCGAGCAGGAATCGG 749

Qy  161 yValTyrTrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGlnThr 181
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   750 CGTTTACTGGGGCCCGGGCCATCCTTTAAATGTAGGCATTAGACTTCCTGGGCGCCAGAC 809

Qy  181 rAsnGlnArgGluIleHisAlaAlaCyLysAlaAlaIleGluGluAlaLysThrGlnAs 201
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   810 AAACCAAAGAGCGGAAATTCATGCAGCCTGCAAAGCCATTGAACAAGCAAAGACTCAAAA 869

Qy  201 nIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsnTr 221
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   870 CATCAATAAACTAGTTCTGTATACAGACAGTATGTTTACGATAAATGTATAACTAACTG 929

Qy  221 pValGlnGlyTyrPLysLysAsnGlyTyrTrpLysThrSerAlaGlyLysGluValIleAlaAsnLy 241
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   930 GGTTCAAGGTTGGAAGAAAAATGGGTGGAAGCAAGTGCAGGGAAAGAGGTGATCAACAA 989

Qy  241 sGluAspPheValAlaAlaLeuArgLeuThrGlnLeuArgLeuMetAspIleGlnTrpMetHisVa 261
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   990 AGAGGACTTTGTTGCACTGAGAGGTTACCCAGGGATGGACATTCAGTGATGCATGT 1049

Qy  261 lProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgGluGlyVal 281
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1050 TCCTGGTCATTCGGGATTTATGGCCATGGAGGAAGGCAATGAAGAGGAAGAGAGC 1109

Qy  281 aLysGlnSerGluAsp 286
        ||||||||||||||
Db  1110 TAAACAATCGGAAGAC 1125


RESULT 8
AF048993
LOCUS       AF048993              1430 bp    mRNA    linear    ROD 21-NOV-1998
DEFINITION  Mus musculus ribonuclease H1 (Rnh1) mRNA, complete cds.
ACCESSION   AF048993
VERSION     AF048993.1  GI:2935417
KEYWORDS    .
```

```
Qy    1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaLeuAlaLeuProCysArgArg  20
          |||  ||| :::  ||
Db   49 ATGAGCTGGCTGCTGCCGCTCCGCGTCGCCCACGTGACACTGGCCGTCGTACGCCTGAGGCGA 108

Qy   21 GlySerArgGlyPheGlyMetPheTyrAlaValArgArgValArgTyrArgLysThrGlyValPhe  40
Db  109 GGCATTTGCGGCTTCGGCATGTTCTATGCGGTGAGGAGAGCCGGAGAGTCTTC 168

Qy   41 LeuThrTrpAsnGluCysArgAlaGlnValAspAspArgPheProAlaAlaArgPheLysLys  60
Db  169 CTGAGTTGGAGTGAGTGCAAAGCCCAGGTGACCGGTTTCCTGCTGCCAGTTTAAGAAA 228

Qy   61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer  80
Db  229 TTTGCCACAGAAGATGAGGCCTGGGCTTTGTCAGGAGCTCTTCAAGCCCGATGGTTCA 288

Qy   81 GluGlyAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100
Db  289 AAAGGGCAGGAAAGTGCACATGACCAGAAGTCACAGCCGAAGACCAGCAAGCGGCCTCGG 348

Qy  101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120
Db  349 GAGCCCCTG---GGTGAAGGGGAAGAACTTCCAGAGCCAGGGCCAAAGCACCACGACAG 405

Qy  121 SerValGluProAlaProProValSerGluArgAspThrPheSerTyrMetGlyAspPheVal 140
Db  406 GACACCGAGCCCGCTGCTGACTGAGCAAGGACACATTTCTTATATGGGAGAGTCAGTC 465

Qy  141 ValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProAlaGlyIle 160
Db  466 ATTGTCTACACGGATGCGTGTTGCTCCAGTAATGACGGAGCGGGCACGAGCAGCAGAATT 525

Qy  161 GlyValTyrTrpGlyTrpGlyHisProLeuAsnValGlyIleArgGlyProGlyTyrArgGln 180
Db  526 GGCGTTTTACTGGGCCCAGGCCACCCCTTAAATGTAGGTATAAGGTTCCTGGGCACAG 585

Qy  181 ThrAsnGlnArgAlaGlyIleHisAlaAlaCysLysLysIleGluGlnAlaLysThrGln 200
Db  586 ACAAACCAGAGGGCGCGAGATCCAGCCATCATGCAAGGCCATCATGCAAGGCCAAGGCTCAG 645

Qy  201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220
Db  646 AACATCAGCAAGCTGGTTCTGTACACAGACAGCATGTTCACCATCAATGGGATAACTAAC 705

Qy  221 TrpValGlnGlyTrpLysAsnGlyTrpLysThrSerAlaGlyValIleLysGluValIleLysAsn 240
Db  706 TGGGTTCAGGGCTGGAAGAATGGCTGGAAGACAAGTACAGGGAAAGATGTGATCAAC 765

Qy  241 LysGluAspPheValAlaLeuGluArgLeuProAspThrGlnGlyMetAspIleGlnTrpMetHis 260
Db  766 AAGGAGGACTTCATGGAGCTGGACGAGCTCACCTCAGGGACATCCAGTGGATGCAC 825

Qy  261 ValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgGluGlyGly 280
Db  826 ATTCCTGGTCACTCAGGATTTGTGGGCAATGAAGAGAGCCGACACTGGCACGCGGGAGGA 885

Qy  281 AlaLysGlnSerGluAsp 286
Db  886 GCGAAGCAGTCTGAGGAC 903
```

```
FEATURES             Location/Qualifiers
     source          1..1430
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="dbEST:AA163387"
                     /db_xref="taxon:10090"
                     /chromosome="12"
                     /map="12A3"
                     /tissue_type="carcinoma"
                     /clone_lib="Stratagene 937317"
                     /dev_stage="embryonic"
     gene            1..1430
                     /gene="Rnh1"
     5'UTR           1..49
                     /gene="Rnh1"
     CDS             49..906
                     /gene="Rnh1"
                     /note="similar to bacterial RNases H1"
                     /codon_start=1
                     /product="ribonuclease H1"
                     /protein_id="AAC78562.1"
                     /db_xref="GI:2935418"
                     /translation="MRWLLPLSRTVTLAVVRLRRGICGLGMFYAVRRGRRTGVFLSWS
                     ECKAQVDRFPAARFKKFATEDEAWAFVRSSSSPDGSKGQESAHEQKSQAKTSKRPREP
                     LGEGEELPEPGPKHTRQDTEPAAVVSKDTFSYMGESVIVYTDGCCSSNGRKRARAGIG
                     VYWGPGHPLNVGIRLPGRQTNQRAEIHAACKAIMQAXAQNISKLVLYTDSMFTINGIT
                     NWVQGWKKNGWRTSTGKDVINKEDFMELDELTQGMDIQWMHIPGHSGFVGNEEADRLA
                     REGAKQSD"
     3'UTR           907..1430
                     /gene="Rnh1"
BASE COUNT        384 a    324 c    414 g    308 t
ORIGIN
```

Alignment Scores:

Label	Value	Label	Value
Pred. No.:	1.06e-77	Length:	1430
Score:	1197.50	Matches:	221
Percent Similarity:	86.71%	Conservative:	27
Best Local Similarity:	77.27%	Mismatches:	37
Query Match:	77.46%	Indels:	1
DB:	10	Gaps:	1

US-10-054-313-1 (1-286) x AF048993 (1-1430)

```
AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
          Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
          Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
          Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
          Diatchenko,L., Matsina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
          Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
          Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
          Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
          Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
          McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
          Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
          Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
          Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
          Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
          Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
          Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
          Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
          Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
TITLE     Generation and initial analysis of more than 15,000 full-length
          human and mouse cDNA sequences
JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
MEDLINE   22388257
PUBMED    12477932
REFERENCE 2  (bases 1 to 1409)
AUTHORS   Strausberg,R.
TITLE     Direct Submission
JOURNAL   Submitted (19-DEC-2001) National Institutes of Health, Mammalian
          Gene Collection (MGC), Cancer Genomics Office, National Cancer
          Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
          USA
REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT   Contact: MGC help desk
          Email: cgapbs-r@mail.nih.gov
          Tissue Procurement: Jeffrey E. Green, M.D.
          cDNA Library Preparation: Life Technologies, Inc.
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Baylor College of Medicine Human Genome
          Sequencing Center
          Center code: BCM-HGSC
          Web site: http://www.hgsc.bcm.tmc.edu/cdna/
          Contact: amg@bcm.tmc.edu
          Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S. W., Loulseged, H.,
          Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
          A.N., Gibbs, R.A.
          Clone distribution: MGC clone distribution information can be found
          through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
          Series: IRAK Plate: 41 Row: n Column: 3
          This clone was selected for full length sequencing because it
          passed the following selection criteria: matched mRNA gi: 6755333.
FEATURES             Location/Qualifiers
     source          1..1409
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="MGC:30220 IMAGE:5133942"
                     /tissue_type="liver, normal. 5 month old male mouse."
                     /clone_lib="NCI_CGAP_Li9"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     gene            1..1409
                     /gene="Rnaseh1"
                     /db_xref="LocusID:19819"
                     /db_xref="MGI:1335073"
     CDS             17..874
                     /codon_start=1
                     /product="ribonuclease H1"
                     /protein_id="AAH19411.1"
                     /db_xref="GI:18044436"
                     /db_xref="LocusID:19819"
                     /translation="MRWLLFLSRTVTLAVVLRRGICGLGMFYAVRGRRTGVFLSWS
                     ECKAQVDRFPAARFKKPATEDEAWAFVRSSSSPDGSKGQSAHEQKSQVKTSKRPREP
```

```
                        LGEGEELPEPGPGHTRQDTEPAAVVSKDTFSYMGESVIVYTDGCCSSNGRKRARAIG
                        VYWGPGHPLNVGIRLPGRQTNQRAEHHACKAIMQAKAQNISKLVLYTDSMFTINGIT
                        NWYQGWKKNGWRTSTGKDVINKEDFMELDELTQGMDIQWMHIPGHSGFVGNEEADRLA
                        REGAKQSED"

BASE COUNT    391 a    315 c    400 g    303 t
ORIGIN

Alignment Scores:
Pred. No.:                1409
Score:                    1193.50
Percent Similarity:       86.36%
Best Local Similarity:    76.92%
Query Match:              77.20%
DB:                       10
```

```
US-10-054-313-1 (1-286)   x BC019411  (1-1409)

Qy    1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaAlaLeuProCysArgArg  20
          |||||||||||||||||||||||||||||||||:|||        :::
Db   17 ATGCGCTGGCTGCTGCTGCTTCCCGACCACAGTGACACTGGCGCCGTCGTACGCCTGAGGCCGA  76

Qy   21 GlySerArgGlyPheGlyMetPheTyrAlaValArgGlyArgArgThrGlyValPhe  40
          ||||  |  ||||||| ::|:|||:|||        :::|||:
Db   77 GGCATTGCGGCTGGGATGTTCATGCCGGTGAGGGGACCGGACCGAGTCTTC  136

Qy   41 LeuThrTrpAsnGluCysArgAlaGlnValAspAspPheProAlaAlaArgPheLysLys  60
          |||:::|| :||||||||:::|||:|:|||:||        :::
Db  137 CTGAGTTGGAGTGAGTGCAAAGCCAGGTGGACCCGTTCCTGCTGCCAGGTTTAAGAAA  196

Qy   61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer  80
          |||||||||||:::|||        |||||:::|||
Db  197 TTTGCCACAGAAGATGAGGCCTGGGCCTTTGTCAGGAGCTCTTCAAGCCCGGATGGTTCA  256

Qy   81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg  100
          |||||:||||||||  |||:||        |||:::|||
Db  257 AAAGGGCAGGAAAGTGCACATGACAGAAGTCACAGTGAAGACCAGCAAGCGGCCTCGG  316

Qy  101 GluProLeuAspGlyGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysProPro  120
          |||||||||    ::|||        :::|||:::|||:::
Db  317 GAGCCCCTG--GGTGAAGGGGAGAACTTCCAGGCCAGGGCCAAAGCACACCACAGCACAG  373

Qy  121 SerValGluProAlaProProValSerArgArgAspThrPheSerTyrMetGlyAspPheVal  140
          |||:||        |:::|||        ||||||:::|||:::|||
Db  374 GACACCGAGCCCGCTGCTGGTAGTGAGCAAGGACACGTTTTCTTATGGGAGTGAGTCAGTC  433

Qy  141 ValValTyrThrAspGlyCysCysSerSerAsnGlyArgLysProAlaGlyAlaGlyIle  160
          :::|||:::|||        :::|||        ||||||:::|||:::|||:::
Db  434 ATTGTCTACACGGATGGCTGTTGCTCCAGTAATGACGGAAGCGGGCACGAGCAGGAATT  493

Qy  161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGln  180
          |||||||||        |||||||||:::|||||||||:::|||:::
Db  494 GGCGTTTACTGGGGCCCAGGCCACCCTAAATGTAAATGTAGGTATAAGGCTTCCTGGGACCAG  553

Qy  181 ThrAsnGlnArgAlaGluHisHisAlaCysLysAlaIleMetGlnAlaLysAlaGlnThrGln  200
          ||||||:::|||        :::|||:::|||:::|||:::
Db  554 ACAAACCAGAGGCCCGAGATCCATGAGCCTGCAAGGCCATCATGCAGGCTAAGGCTCAG  613

Qy  201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn  220
          |||:::|||        :::|||:::|||:::|||:::|||:::|||:::
Db  614 AACATCAACAAGCTGGTTCTGTACACAGACAGCATGTTCACCATCAATGGGATAACTAAC  673

Qy  221 TrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysAspValIleAsn  240
          |||||||||        :::|||:::|||:::|||:::
Db  674 TGGGTTCAGGGCTGGAAGAAGAATGGCTGGAAGACAAGTACAGGGAAAGAGTGTGATCAAC  733

Qy  241 LysGluAspPheValAlaLeuGluArgLeuThrGlnLeuThrMetAspIleGlnTrpMetHis  260
          |||:::|||        :::|||:::|||:::|||:::|||:::
Db  734 AAGGAGGACTTCATGGAGCTGGACGAGCTCACTCAGGGCATGGACATCCAATGGATGCAC  793

Qy  261 ValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgGluGly  280
          :::|||        |||:::|||:::|||:::|||:::
Db  794 ATTCCTGGTCACTCAGGATTTGTGGGCAATGAAGAGGCCGACAGACTGGCACGGGAAGGA  853
```

```
Qy          281 AlaLysGlnSerGluAsp 286
                ||||||||||||||||||
Db          854 GCGAAGCAGTCTGAGGAC 871

RESULT 10
AC098850
LOCUS       AC098850            178965 bp    DNA     linear   PRI 30-DEC-2001
DEFINITION  Homo sapiens chromosome 17, clone RP11-92B11, complete sequence.
ACCESSION   AC098850
VERSION     AC098850.2  GI:17646937
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 178965)
AUTHORS     Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE       Homo sapiens chromosome 17, clone RP11-92B11
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 178965)
AUTHORS     Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McSwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE       Direct Submission
JOURNAL     Submitted (03-NOV-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 178965)
AUTHORS     Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McSwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Zimmer,A. and Zody,M.
TITLE       Direct Submission
JOURNAL     Submitted (30-DEC-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Dec 13, 2001 this sequence version replaced gi:16604041.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            -------- Project Information
            Center project name: L13829
            Center clone name: 92_B_11
            --------

FEATURES             Location/Qualifiers
     source          1..178965
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="RP11-92B11"
                     /clone_lib="RPCI-11 Human Male BAC"
     unsure          987..1002
                     /note="<30 qual SNGL region"
     repeat_region   994..1308
                     /rpt_family="AluJo"
     repeat_region   complement(1434..1735)
                     /rpt_family="AluSx"
     repeat_region   2075..2095
                     /rpt_family="AT_rich"
     repeat_region   complement(2205..2509)
                     /rpt_family="AluSx"
     repeat_region   2558..2592
                     /rpt_family="AT_rich"
     repeat_region   2730..3033
                     /rpt_family="AluJo"
     repeat_region   3034..3339
                     /rpt_family="AluSx"
     repeat_region   4237..4520
                     /rpt_family="AluSx"
     repeat_region   4584..5144
                     /rpt_family="L1ME"
     repeat_region   5180..5481
                     /rpt_family="AluSx"
     repeat_region   complement(5496..5542)
                     /rpt_family="AluSx"
     repeat_region   complement(5543..5678)
                     /rpt_family="AluSx"
     repeat_region   complement(5801..6120)
                     /rpt_family="AluSx"
     repeat_region   6123..6363
                     /rpt_family="MER58B"
     repeat_region   6423..6533
                     /rpt_family="L1ME"
     repeat_region   6534..6830
                     /rpt_family="L1ME"
     repeat_region   6831..7175
                     /rpt_family="AluSx"
     repeat_region   7310..7435
                     /rpt_family="L1ME"
     repeat_region   7527..7703
                     /rpt_family="L1ME"
     repeat_region   7704..7993
                     /rpt_family="L1ME"
     repeat_region   7994..8309
                     /rpt_family="AluJo"
     repeat_region   8319..8807
                     /rpt_family="L1ME"
     repeat_region   8893..9579
                     /rpt_family="L2"
     repeat_region   9941..9994
                     /rpt_family="GA-rich"
     repeat_region   complement(10332..10491)
                     /rpt_family="MER3"
     repeat_region   complement(11158..11392)
```

```
Db  59696  ATGAGCCGGCTTCTGTCCTGGCGGC-AGAGTCGGCTCCCTTGGCGCCGCCTTGCCTGCCGCCGC  59754
Qy     21  GlySerArgGlyMetPheTyrAlaValArgArgGlyValHisArgGlyAlaGlyThrGlyLeuValPhe  40
Db  59755  ---TCTCGCGGGTTCGGGATGTTCATGCCGTGAGGAGGGGCGGCAAGTCCGCGGTCTTT  59811
Qy     41  LeuThrTrpAsnGlyCysArgAlaGlnValAlaAspArgPheProAlaAlaArgPheLysLys  60
Db  59812  CTGACCCGGAATGAGTGCAAAGCACAGGTGACCCGGTTCTGCTGCCAGATTTAAGAAG  59871
Qy     61  PheAlaThrGluAspGluAlaTrpAlaPheValLysSerAlaSerProGluValSer  80
Db  59872  TTTGCCACAGAGGATGAGGCCTGGGACTTTGTCAGGAAATCTGCAAGCCCGGAAGTTTCA  59931
Qy     81  GluGlyHisGluAsnGlnHisGlyGlnGluSerAlaLysProGlyLysArgLeuArg  100
Db  59932  GAAGGGCAGGAAAATCAACATGGACAAGAATCGGACAAGAACGCAGCAAGCAGCGACTCCGT  59991
Qy    101  GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMet------  118
Db  59992  GAGCCACTGGATGGAGATGGACATATGC-AGCCACATGAAGCCG  60050
Qy    119  ------LysProSerValGluProAlaProValSerArgArgAspThr  132
Db  60051  AGCGTGAAGCCGAGCGTGAAGCCGCCTCCAGTTAGCAGAGACACG  60110
Qy    133  PheSerTyrMetGlyAspPheValValValTyrThrAspGlyCysCysSerSerAsnGly  152
Db  60111  TTTTCCTACATGGGAGACTTCGTCGTCGTCACGCTGATGCGTCGTCCAGTAATGGG  60170
Qy    153  ArgArgLysProArgAlaGlyIleGlyValTyrArgGlyValHisProLeuAsnVal  172
Db  60171  CGTAGAAAGCCACGAGCAGGAATCCGGCTTTACTGGGGGCCGGGGCTATCCTTTAACTGTA  60230
Qy    173  GlyIleArgLeuProGlyArgGlnThrAsnGlnArgAlaGluIleHisAlaAlaCysLys  192
Db  60231  GGCATTAGACTTCCTGGGTGGCAGACAAACCAAAGAGCGAAATTCATGCAGCCTGCAAG  60290
Qy    193  AlaIleGluGlnAlaLysThrGlnAlaLysIleAsnIleAsnLysValLeuTyrThrAspSerMet  212
Db  60291  GCCATTGAACAGGCCAAGACTCAGAAAATCAATAAACTCGGTTCTGTATACACAGACAGTATG  60350
Qy    213  PheThrIleAsnGlyIleThrAsnTrpValGlnValTrpLysAsnGlyThrTrpLysSerThr  232
Db  60351  TTTACGATAAATGGTATAACTAACTGGGTTCAAGTTGGAAGAAAAATGGGTGGAAGACA  60410
Qy    233  SerAlaGlyLysValIleIleAsnLysGluAsnLysGluValAlaAlaLeuGluLeuArgGlyLeuArgLeuThrGln  252
Db  60411  AGTGCAGGGAAAGAGGTGATCAACAAGAGAGACTTCGTGGCAACTGAGGAGGCTGACCCCAG  60470
Qy    253  GlyMetAspIleGlnTrp-------------MetHisValPro  262
Db  60471  GGGATGACATTCAGTGGGCAGTATCTTAAATGTTCATGTCCCA  60515
```

```
repeat_region    /rpt_family="L3"
                 11639..12721
repeat_region    /rpt_family="LiMC5"
                 12814..13108
repeat_region    /rpt_family="AluY"
                 complement(13131..13216)
repeat_region    /rpt_family="L1P"
                 13203..13257
repeat_region    /rpt_family="L1PA4"
                 13255..15525
repeat_region    /rpt_family="L1PA4"
                 complement(15190..15383)
repeat_region    /rpt_family="MIR"
                 15752..16394
repeat_region    /rpt_family="LiMB7"
                 16392..16515
repeat_region    /rpt_family="LiMB7"
                 16515..17530
repeat_region    /rpt_family="LiMB7"
                 complement(17531..17840)
repeat_region    /rpt_family="AluSx"
                 17841..19946
repeat_region    /rpt_family="LiMB7"
                 20113..20529
repeat_region    /rpt_family="MSTB1"
                 20934..21416
repeat_region    /rpt_family="LTR65"
                 21417..21594
repeat_region    /rpt_family="AluSq"
                 21929..22229
repeat_region    /rpt_family="AluSx"
                 complement(22689..23156)
repeat_region    /rpt_family="LTR37B"
                 complement(24183..24502)
repeat_region    /rpt_family="AluSx"
                 complement(26205..26386)
repeat_region    /rpt_family="MIR"
                 26671..26697
repeat_region    /rpt_family="(TAA)n"
                 27404..27705
repeat_region    /rpt_family="AluSx"
                 complement(28018..28146)
repeat_region    /rpt_family="MER5A"
                 28157..28214
repeat_region    /rpt_family="AT_rich"
                 28714..28743
repeat_region    /rpt_family="AT_rich"
                 28825..28870
repeat_region    /rpt_family="AT_rich"
                 31073..31292
repeat_region    /rpt_family="LiME4A"
                 31553..31858
repeat_region    /rpt_family="L3"
                 32320..33351
repeat_region    /rpt_family="AT_rich"
                 32720..33020
repeat_region    /rpt_family="AluJo"
                 33053..33370
repeat_region    /rpt_family="AluYb8"
                 complement(33693..34052)
```

```
Alignment Scores:
Pred. No.:
Score:                  1.13e-73        Length:         178965
                        1174.00         Matches:        236
Percent Similarity:     88.36%          Conservative:   7
Best Local Similarity:  85.82%          Mismatches:     18
Query Match:            75.94%          Indels:         16
DB:                     9               Gaps:           3

US-10-054-313-1 (1-286) x AC098850 (1-178965)

Qy   1  MetSerTrpLeuPheLeuAlaHisArgValAlaLeuAlaAlaLeuAlaLeuProCysArgArg  20
```

```
          Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
          DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
          Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
          Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
          Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
          Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
          Macdonald,P., Marquis,N., McGurk,A., McKernan,K.,
          McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
          Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
          Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
          Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
          Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
          Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
          Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3 (bases 1 to 192337)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Beda,F., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,
            FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,
            Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L.,
            Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G.,
            Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,
            O'Donnell,P., O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Rieback,M., Riley,R.,
            Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-SEP-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
  COMMENT   On Sep 23, 2000 this sequence version replaced gi:9795639.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ------------ Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence submissions@genome.wi.mit.edu
            ------------ Project Information
            Center project name: L4191
            Center clone name: 219_A_15

FEATURES             Location/Qualifiers
     source          1..192337
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="RP11-219A15"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   568..788
                     /rpt_family="MLT1C"
     repeat_region   1804..1957
                     /rpt_family="FRAM"
     repeat_region   1964..2014
                     /rpt_family="MIR"
     repeat_region   complement(2029..2319)
                     /rpt_family="LTR16C"
     repeat_region   2463..2535
                     /rpt_family="MIR"
```

```
     repeat_region   complement(2614..2700)
                     /rpt_family="MIR3"
     repeat_region   complement(2938..3185)
                     /rpt_family="L1MC/D"
     repeat_region   complement(4802..4869)
                     /rpt_family="L2"
     repeat_region   complement(4871..5058)
                     /rpt_family="LTR16B"
     repeat_region   complement(5150..5227)
                     /rpt_family="LTR16A"
     repeat_region   5353..5597
                     /rpt_family="AluJo"
     repeat_region   5697..5863
                     /rpt_family="MLT1J"
     repeat_region   complement(5909..6210)
                     /rpt_family="AluSx"
     repeat_region   complement(6235..6609)
                     /rpt_family="L1MB3"
     repeat_region   6610..6687
                     /rpt_family="AluSg/x"
     repeat_region   complement(6688..6802)
                     /rpt_family="L1MB3"
     repeat_region   7399..7751
                     /rpt_family="MSTD"
     repeat_region   7977..8102
                     /rpt_family="MER5A"
     repeat_region   complement(8265..8596)
                     /rpt_family="L1ME3"
     repeat_region   complement(10173..10350)
                     /rpt_family="MIR"
     repeat_region   complement(10370..10452)
                     /rpt_family="MIR"
     repeat_region   10648..11010
                     /rpt_family="L1ME4A"
     repeat_region   complement(11168..11273)
                     /rpt_family="L1MC4a"
     repeat_region   complement(11274..11583)
                     /rpt_family="AluSx"
     repeat_region   complement(11584..11603)
                     /rpt_family="L1MC4"
     repeat_region   12176..12289
                     /rpt_family="MIR"
     repeat_region   complement(12402..12464)
                     /rpt_family="L2"
     repeat_region   13177..13351
                     /rpt_family="MER69"
     repeat_region   13950..14083
                     /rpt_family="MIR"
     repeat_region   complement(14660..15282)
                     /rpt_family="L1MC4a"
     repeat_region   complement(15834..15968)
                     /rpt_family="L1MC4"
     repeat_region   complement(16176..16237)
                     /rpt_family="L2"
     repeat_region   17027..17098
                     /rpt_family="MIR"
     repeat_region   complement(17196..17328)
                     /rpt_family="L2"
     repeat_region   complement(17390..17492)
                     /rpt_family="L2"
     repeat_region   18087..18429
                     /rpt_family="L2"
     repeat_region   18524..18675
                     /rpt_family="MER3"
     repeat_region   18848..19158
                     /rpt_family="AluSp"
     repeat_region   complement(19554..19800)
                     /rpt_family="AluSx"
     repeat_region   complement(20989..21284)
                     /rpt_family="AluY"
     repeat_region   21286..21365
                     /rpt_family="MER3"
     repeat_region   21393..21576
```

```
Db   191206  AGGGTGAAGCCGAGCGTGAAGCCGCGTGAAGCGCGGCGGCCGCCTGAAGCGCGGCGGCCCTCCAGTTAGCAGAGACACG  191147
                   |||||||||||||||||||||||||||||||||||||||||||||||||||::::||||||||||||||||||||||
Qy      133  PheSerTyrMetGlyAspPheValValValTyrThrAspGlyCysCysSerSerAsnGly  152
Db   191146  TTTTCCTACATGGGAGACTTCGTCGTCGTCACGGTGGCTCGCTGCTCCCAGTAATGGG  191087

Qy      153  ArgArgLysProArgAlaGlyIleGlyValTyrTrpProGlyHisProLeuAsnVal  172
                   ||||||||||||||||||||||||||||||||::::|||||||
Db   191086  CGTAGAAAGCCACGAGCAGGAATCCGGGTTACTGGGGCCGGGCTATCCTTTAACTGTA  191027

Qy      173  GlyIleArgLeuProGlyArgGlnThrAsnGlnArgAlaGlnArgAlaGlyIleHisAlaAlaCysLys  192
                   |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   191026  GGCATTAGACTTCCTGGGTGGCAGACAAACCAAAGAGCGAAATCATGCAGCCTGCAAG  190967

Qy      193  AlaIleGluGlnAlaLysThrGlnAsnIleAsnLysLeuValLeuTyrThrAspSerMet  212
                   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   190966  GCCATTGAACAAGCACAGACTCAGAAAATCAATAAACTGGTTCTGTATACAGACAGTATG  190907

Qy      213  PheThrIleAsnGlyIleThrArgAsnTrpValGlnIleGlyTrpLysLysAsnGlyTrpLysThr  232
                   |||||||||||||||||||||||||||||||||||||||||||||||||
Db   190906  TTTACGATAAATGGTATAACTAACTGGGTTCAAGGTTGGAAGAAAAATGGGTGGAAGACA  190847

Qy      233  SerAlaGlyLysGluValIleLeuAsnLysGluAspPheValAlaGluGlyArgLeuThrGln  252
                   |||||||||||||||||||||||||||||||||||||||||||||||||
Db   190846  AGTGCAGGGAAAGAGGTGATCAACAAAGAGGACTTCGTGGCACTGGAGAGGCTGACCCAG  190787

Qy      253  GlyMetAspIleGlnTrp----------MetHisValPro  262
                   |||||||||||||||              :::||||||||
Db   190786  GGGATGGACATTCAGTGGGCGGAGTATCTTAAATGTTCATGTCCCA  190742
```

```
RESULT 12
AC126352/c
LOCUS        AC126352            211403 bp    DNA        linear   PRI 19-MAR-2003
DEFINITION   Homo sapiens chromosome 17, clone CTD-2509K7, complete sequence.
ACCESSION    AC126352
VERSION      AC126352.6  GI:29124211
KEYWORDS     HTG.
SOURCE       Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 211403)
  AUTHORS    Birren,B., Nusbaum,C. and Lander,E.
  TITLE      Homo sapiens chromosome 17, clone CTD-2509K7
  JOURNAL    Unpublished
REFERENCE    2  (bases 1 to 211403)
  AUTHORS    Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
             Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
             Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
             Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
             Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
             Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
             Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
             Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
             Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
             McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
             Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
             O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,K., Peterson,K.,
             Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
             Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
             Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
             Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
             Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
             Zembek,L., Zimmer,A. and Zody,M.
  TITLE      Direct Submission
  JOURNAL    Submitted (05-JUL-2002) Whitehead Institute/MIT Center for Genome
             Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE    3  (bases 1 to 211403)
  AUTHORS    Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
             Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
             Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
             Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
```

```
repeat_region   /rpt_family="L2"
                21627..21701
repeat_region   /rpt_family="MIR"
                21770..22001
repeat_region   /rpt_family="LTR14A"
                complement(22014..22401)
repeat_region   /rpt_family="L1MA4A"
                23317..23339
repeat_region   /rpt_family="AT_rich"
                complement(23367..23502)
repeat_region   /rpt_family="MIR"
                24118..24141
repeat_region   /rpt_family="(TCTG)n"
                complement(24473..24616)
repeat_region   /rpt_family="MIR3"
                25529..28376
repeat_region   /rpt_family="L2"
                complement(25652..26053)
repeat_region   /rpt_family="MIR3"
                2829..28376
repeat_region   /rpt_family="MIR"
                34302..34408
repeat_region   /rpt_family="MIR"
                complement(35693..35781)
repeat_region   /rpt_family="MIR"
                37220..37405
repeat_region   /rpt_family="L2"
                complement(37431..37554)
repeat_region   /rpt_family="MIR"
                38931..39908
repeat_region   /rpt_family="LTR5"
                40343..40448
repeat_region   /rpt_family="MIR"
                40552..40629
repeat_region   /rpt_family="MIR"
                complement(41149..41275)
repeat_region   /rpt_family="MIR"
                complement(41355..41506)
                /rpt_family="MIR3"
```

```
Alignment Scores:
Pred. No.:
Score:              1.22e-73
Percent Similarity: 1174.00
Best Local Similarity: 88.36%
Query Match:         85.82%
DB:                  75.94%
                     9
Length:    192337
Matches:   236
Conservative: 7
Mismatches:  18
Indels:    16
Gaps:      3

US-10-054-313-1 (1-286) x AC022596 (1-192337)

Qy      1  MetSerTrpLeuPheLeuAlaHisArgValAlaAlaLeuAlaAlaLeuProCysArgArg  20
                 |||||||||||||||||||||||||||||||||||||||||
Db  191561  ATGAGCTGGCTTCTGTCCCTGGCGGC-AGAGTCGCCTTGGCCGCCGCCTTGCCCGCCGCC  191503

Qy     21  GlySerArgGlyPheGlyMetPheTyrAlaValArgArgGlyArgGlyLysThrGlyValPhe  40
                 ||||||||||||||||||||||||||||||||||::::|||||
Db  191502  ---TCTCGCGGGTTCGGGATGTTCTATGCCGTGAGAGGGCGGCAGTCCGCGGTCTTT  191446

Qy     41  LeuThrTrpAsnGluCysArgAlaGlnValAspArgArgPheProAlaAlaArgPheLysLys  60
                 |||||||||||:::|||||||||||||||||||
Db  191445  CTGACCCGGGAATGAGTGCAAAGCACAGGTGGACCGGGTTCCTGCTGCCAGATTTAAGAAG  191386

Qy     61  PheAlaThrGluAspGluArgValAlaPheValArgSerAlaSerProGluValSer  80
                 ||||||||||||||||||||||||||||||||||
Db  191385  TTTGCCACAGAGGATGAGCCCTGGGACTTTGTCAGGAAATCTGCAAGCCCGGAAGTTTCA  191326

Qy     81  GluGlyHisGluAsnGlnHisGlyGlnGluSerGluGluAlaLysProGlyLyLysProGlyLeuArg  100
                 ||||||||::::|||
Db  191325  GAAGGGCAGGAAAATCAACATGGACAAAGAAATCCGAGACAAGAAGCCAGCAGGCGACTCCGT  191266

Qy    101  GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMet------  118
                 ||||||||||||::::||||||||||||
Db  191265  GAGCCACTGGATGGAGATGGAAAGCGCACATATGC-AAGCACATGAAGCCG  191207

Qy    119  --------LysProSerValGluProValAlaProProValSerProArgAspThr  132
```

```
                 Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
                 Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
                 Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
                 Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
                 Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J.,
                 Matthews,C., McCarthy,M., Meldrin,J., Meneus,J., Mihova,T.,
                 Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
                 Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
                 Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
                 Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,C., Schupback,R.,
                 Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
                 Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
                 Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
                 Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE            Direct Submission
JOURNAL          Submitted (20-NOV-2002) Whitehead Institute/MIT Center for Genome
                 Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE        4  (bases 1 to 211403)
AUTHORS          Birren,B., Nusbaum.C., Lander,E., Abouelleil,A., Allen,N.,
                 Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
                 Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
                 Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
                 Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
                 Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
                 Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
                 Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
                 Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
                 Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
                 Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
                 Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
                 Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
                 O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
                 Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
                 Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
                 Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
                 Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
                 Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
                 Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE            Direct Submission
JOURNAL          Submitted (19-MAR-2003) Whitehead Institute/MIT Center for Genome
                 Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT          On Mar 19, 2003 this sequence version replaced gi:25140067.
                 All repeats were identified using RepeatMasker:
                 Smit, A.F.A. & Green, P. (1996-1997)
                 http://ftp.genome.washington.edu/RM/RepeatMasker.html
                 --------------- Genome Center
                    Center: Whitehead Institute/ MIT Center for Genome Research
                    Center code: WIBR
                    Web site: http://www-seq.wi.mit.edu
                    Contact: sequence submissions@genome.wi.mit.edu
                 ---------- Project Information
                    Center project name: L27684
                    Center clone name: 2509_K_7
                 ----------
FEATURES             Location/Qualifiers
     source          1..211403
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="CTD-2509K7"
                     /clone_lib="CITD2 Human BAC"
     repeat_region   3..67
                     /rpt_family="MER5A"
     repeat_region   complement(230..561)
                     /rpt_family="L1ME3"
     repeat_region   1679..1776
                     /rpt_family="L3b"
     repeat_region   complement(2138..2325)
                     /rpt_family="MIR"
     repeat_region   complement(2335..2417)
                     /rpt_family="MIR"
```

```
     repeat_region   2613..2966
                     /rpt_family="L1MB4a"
     repeat_region   complement(3133..3237)
                     /rpt_family="L1MC4"
     repeat_region   complement(3239..3547)
                     /rpt_family="AluSx"
     repeat_region   4140..4268
                     /rpt_family="MIR"
     repeat_region   complement(4365..4427)
                     /rpt_family="L2"
     repeat_region   5140..5314
                     /rpt_family="MER69"
     repeat_region   5913..6046
                     /rpt_family="MIR"
     repeat_region   complement(6623..7245)
                     /rpt_family="L1MC4a"
     repeat_region   complement(7797..7931)
                     /rpt_family="MIR3"
     repeat_region   complement(8139..8200)
                     /rpt_family="L2"
     repeat_region   8990..9061
                     /rpt_family="MIR"
     repeat_region   complement(9159..9291)
                     /rpt_family="L2"
     repeat_region   complement(9353..9455)
                     /rpt_family="L2"
     repeat_region   complement(9570..9638)
                     /rpt_family="L3b"
     repeat_region   10050..10392
                     /rpt_family="L2"
     repeat_region   10487..10638
                     /rpt_family="MER3"
     repeat_region   10811..11121
                     /rpt_family="AluSp"
     repeat_region   complement(11517..11763)
                     /rpt_family="AluSx"
     unsure          11534..11544
                     /note="<30 qual SNGL region"
     unsure          11554..11592
                     /note="<30 qual SNGL region"
     repeat_region   complement(12952..13247)
                     /rpt_family="AluY"
     repeat_region   13249..13328
                     /rpt_family="MER3"
     repeat_region   13356..13539
                     /rpt_family="L2"
     repeat_region   13590..13664
                     /rpt_family="MIR"
     repeat_region   13733..13964
                     /rpt_family="LTR14A"
     repeat_region   complement(13965..14364)
                     /rpt_family="L1PB4"
     repeat_region   15280..15302
                     /rpt_family="AT_rich"
     repeat_region   complement(15330..15465)
                     /rpt_family="MIR"
     repeat_region   16082..16105
                     /rpt_family="(TCTG)n"
     repeat_region   complement(16437..16580)
                     /rpt_family="MIR3"
     repeat_region   complement(17671..18017)
                     /rpt_family="L2"
     repeat_region   complement(18266..18369)
                     /rpt_family="MIR"
     repeat_region   20257..20340
                     /rpt_family="MIR3"
     repeat_region   26262..26368
                     /rpt_family="MIR"
     repeat_region   complement(27651..27739)
                     /rpt_family="MIR"
     repeat_region   29178..29363
                     /rpt_family="L2"
     repeat_region   complement(29389..29512)
```

```
repeat_region    /rpt_family="MIR"
                 30888..31865
repeat_region    /rpt_family="LTR5_Hs"
                 32286..32405
repeat_region    /rpt_family="MIR"
                 32509..32586
repeat_region    /rpt_family="MIR"
                 complement(33106..33232)
repeat_region    /rpt_family="MIR"
                 complement(33312..33463)
repeat_region    /rpt_family="MIR3"
                 34251..34325
repeat_region    /rpt_family="(TGG)n"
                 complement(35953..36046)

Alignment Scores:
Pred. No.:                           211403
Score:                  6.06e-73     235
Percent Similarity:     1155.00      7
Best Local Similarity:  88.00%       19
Query Match:            85.45%       16
DB:                     75.36%       3
                        9
```

```
Qy      233 SerAlaGlyLysGluValIleAsnLysGluArgLeuArgThrGln 252
            ||||||||||||||||||||||||||||||||||||||||||||||
Db   182747 AGTGCAGGGAAAGAGGTGATCAACAAAGAGACTTCGTGGCACTGAGAGAGGCTGACCCAG 182688

Qy      253 GlyMetAspIleGlnTrp------------MetHisValPro 262
            |||||||||||||||                :::|||||||||
Db   182687 GGGATGACATTCAGTGGCCAGTATCTTAAATGTTCATGTCCCA 182643
```

```
RESULT 13
AC090774
LOCUS        AC090774              207418 bp    DNA     linear   PRI 28-JUN-2002
DEFINITION   Homo sapiens chromosome 17, clone RP11-344E13, complete sequence.
ACCESSION    AC090774
VERSION      AC090774.6  GI:21617728
KEYWORDS     HTG.
SOURCE       Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 207418)
  AUTHORS    Birren,B., Nusbaum,C. and Lander,E.
  TITLE      Unpublished
REFERENCE    2  (bases 1 to 207418)
  AUTHORS    Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
             Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
             Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
             Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
             Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
             Gardyna,S., Ginde,S., Goyette,P., Graham,L., Grand-Pierre,N.,
             Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
             Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
             Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
             Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
             McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
             Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
             O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
             Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
             Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
             Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
             Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
             Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
             Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
             Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zaincun,J.,
             Zembek,L., Zimmer,A. and Zody,M.
  TITLE      Direct Submission
  JOURNAL    Submitted (10-MAR-2001) Whitehead Institute/MIT Center for Genome
             Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE    3  (bases 1 to 207418)
  AUTHORS    Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
             Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
             Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
             Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
             Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
             Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
             Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
             Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
             Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
             McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
             Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
             O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
             Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
             Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
             Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
             Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
             Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zaincun,J.,
             Zembek,L., Zimmer,A. and Zody,M.
  TITLE      Direct Submission
  JOURNAL    Submitted (27-JUN-2002) Whitehead Institute/MIT Center for Genome
             Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE    4  (bases 1 to 207418)
  AUTHORS    Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
             Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
```

```
US-10-054-313-1 (1-286) x AC126352 (1-211403)

Qy        1 MetSerTrpLeuPheLeuHisArgValAlaLeuAlaAlaLeuProCysArgArg   20
            |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   183462 ATGAGCCGGCTTCTGTCCCTGGCGGC-AGAGTCCCCTTGGCCCCTGCCCTGCCCGCGC 183404

Qy       21 GlySerArgGlyMetPhePheTyrAlaValArgArgGlyArgLysThrGlyValPhe   40
            |||||||||||||||| ||||||| || ||||||||||| |||::||||||||||
Db   183403 ---TCTCGGCGGGTTCGGGATGTTCTATGCCGTGAGGAGGGGGCGCAAGTCCCGGTCTTT 183347

Qy       41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPhePrGluAlaArgPheLysLys   60
            |||||||::||||| ||||||||  |||||||||||  ||||||||||||||||||
Db   183346 CTGACCCGGAATGAGTGCAAGGCACAGGTGGACCGGTTTCCTGCTGCCAGATTTAAGAAG 183287

Qy       61 PheAlaThrGluAspGluAlaTrpAlaPhePheValArgLysSerAlaSerProGluValSer   80
            |||||||||||||||||||||||||||||||||||||| ||||||| |||||||||
Db   183286 TTTGCCACAGAGGATGAGCCTGGGACTTTGTCAGGAAATCTGCAAGCCCGGAAGTTTCA 183227

Qy       81 GluGlyHisGluAsnGlnHisGlyTyrGlnGluSerGluAlaAlaLysProGlyLysArgLeuArg   100
            |||||||||||||||||||:|||| || |||||||||||||||||||||||||||
Db   183226 GAAGGGCAGGAGAAATCAACATGACAAGAATCGGAGACAAGAAGCCAGCAAGCGACTCCGT 183167

Qy      101 GluProLeuAspGlyAspHisGlySerAlaGlnProTyrAlaLysHisMet------   118
            ||||||||||||||||||||||||||||||||  ||||||||||||||||||
Db   183166 GAGCCACTGGATGGAGATGGAGATGGAGAAACGCAGCCCATATGC-AAGCACATGAAGCCG 183108

Qy      119 ------------LysProSerValGluProAlaProProValSerArgArgAspThr   132
                        |:||:|| ||||||||||||||||||||||||||||||||
Db   183107 AGTGTGAAGCCGAGCGTGAAGCCGGCGCCTCCAGTTAGCAGAGACACG 183048

Qy      133 PheSerTyrMetGlyAspPheValValTyrThrAspGlyCysCysSerSerAsnGly   152
            |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   183047 TTTTCCTACATGGGAGACTTCGTCGTCTACACGGATGGCTGCTGCTCCAGTAATGGG 182988

Qy      153 ArgArgLysProArgAlaGlyLysValValTyrTrpGlyValTyrTyrProGlyHisProLeuAsnVal   172
            |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   182987 CGTAGAAAGCCACGAGCAGGAATCCCGCGTTCAGGTGGTGTACTGGGGCGTGTACTATCCTTTAAACTGTA 182928

Qy      173 GlyIleArgLeuProGlyArgGlnThrArgAsnGlnArgAlaGluIleHisAlaAlaCysLys   192
            |||||||||||||||||| ||||||||:|||||||||||||||||||||||||||
Db   182927 GGCATTAGACTTCCTGGGTGGCAGACAATCCAAAGAGCCGAAATTCATGCAGCCTGCAAG 182868

Qy      193 AlaIleGluGlnAlaLysThrGlnAsnIleAsnLysLeuValLeuTyrThrAspSerMet   212
            |||||||||||||:||:||||||||||||||||||||||||||||||||||||
Db   182867 GCCATTGAACAAGCAAGGACTCAGAAAATCAATAAACTGGTTCGTATACAGACAGTATG 182808

Qy      213 PheThrIleAsnGlyIleThrAsnTrpValGlnGlyValTrpLysLysAsnGlyTrpLysTrpThr   232
            |||||:|::||||||||||||||||||||||||||||||||||||||||||||
Db   182807 TTTACGATAAATGGTATAACTAACTGGGTTCAAGGTTGGAAGAAAAATGGGTGGAAGACA 182748
```

```
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
TITLE       Direct Submission
JOURNAL     Submitted (28-JUN-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jun 27, 2002 this sequence version replaced gi:21426207.
            All repeats were identified using RepeatMasker:
            Smit,A.F.A. & Green,P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ----- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            ----- Project Information
            Center project name: L12729
            Center clone name: 344_E_13
            -----
FEATURES             Location/Qualifiers
source          1..207418
                /organism="Homo sapiens"
                /mol_type="genomic DNA"
                /db_xref="taxon:9606"
                /chromosome="17"
                /map="17"
                /clone="RP11-344E13"
                /clone_lib="RPCI-11 Human Male BAC"
repeat_region   1..264
                /rpt_family="LTR35"
repeat_region   265..499
                /rpt_family="L1PA13"
repeat_region   949..1259
                /rpt_family="AluY"
repeat_region   1268..1401
                /rpt_family="L1MC/D"
repeat_region   2066..2204
                /rpt_family="LTR67"
repeat_region   2210..2340
                /rpt_family="L1MC/D"
repeat_region   complement(2412..2546)
                /rpt_family="L2"
repeat_region   complement(3163..3459)
                /rpt_family="AluSq"
repeat_region   6492..6534
                /rpt_family="(TG)n"
repeat_region   6753..6808
                /rpt_family="AT_rich"
repeat_region   complement(7108..7407)
                /rpt_family="AluY"
repeat_region   7825..8108
                /rpt_family="AluY"
repeat_region   complement(8512..8547)
                /rpt_family="MADE1"
repeat_region   8875..8940
                /rpt_family="L2"
repeat_region   9921..10228
                /rpt_family="AluSx"
repeat_region   10713..11010
                /rpt_family="AluJb"

repeat_region   12697..12804
                /rpt_family="FLAM_A"
repeat_region   complement(12850..13549)
                /rpt_family="L1PA10"
repeat_region   13550..13850
                /rpt_family="AluSp"
repeat_region   complement(13851..14554)
                /rpt_family="L1PA10"
repeat_region   14680..15680
                /rpt_family="L1PA15"
repeat_region   15862..15967
                /rpt_family="L2"
repeat_region   15968..15988
                /rpt_family="(TGAA)n"
repeat_region   16061..17124
                /rpt_family="L1M4"
repeat_region   17590..17664
                /rpt_family="L1M4"
repeat_region   17665..17967
                /rpt_family="AluY"
repeat_region   17968..17979
                /rpt_family="L1M4"
repeat_region   17980..18006
                /rpt_family="(CAAA)n"
repeat_region   18007..18381
                /rpt_family="L1M4"
repeat_region   20417..20580
                /rpt_family="AluY"
repeat_region   20586..21111
                /rpt_family="L1M4"
repeat_region   complement(21112..21403)
                /rpt_family="AluSx"
repeat_region   21404..22015
                /rpt_family="L1M4"
repeat_region   22063..22094
                /rpt_family="AT_rich"
repeat_region   23322..23380
                /rpt_family="AluY"
repeat_region   complement(23381..23670)
                /rpt_family="(TA)n"
repeat_region   complement(25468..26047)
                /rpt_family="AluSq"
repeat_region   complement(26143..26832)
                /rpt_family="L1ME3A"
repeat_region   27610..27727
                /rpt_family="L1ME3A"
repeat_region   27728..28199
                /rpt_family="AluY"
repeat_region   28200..28408
                /rpt_family="L1PA4"
repeat_region   28936..29315
                /rpt_family="AluY"
repeat_region   30962..31259
                /rpt_family="MSTD"
repeat_region   complement(29697..30168)
                /rpt_family="L1MC2"
repeat_region   30281..30372
                /rpt_family="(TA)n"
repeat_region   complement(30403..30510)
                /rpt_family="MIR"
repeat_region   complement(31773..32400)
                /rpt_family="L2"
repeat_region   32401..32428
                /rpt_family="AT_rich"

Alignment Scores:
Pred. No.:                  6.61e-72      Length:        207418
Score:              1150.50              Matches:       234
Percent Similarity: 89.89%              Conservative:  6
Best Local Similarity: 87.64%          Mismatches:    20
Query Match:        74.42%              Indels:        10
DB:                 9                    Gaps:          3
```

US-10-054-313-1 (1-286) x AC090774 (1-207418)

```
Qy      1 MetSerTrpLeuPheLeuAlaHisArgValAlaLeuAlaAlaLeuProCysArgArg 20
          |||||||||||||||||||||||||||||||||||||||||||||      |||
Db  58364 ATGAGCTGGCTGTTCTGTTCCTGGCCGC-AGAGCGCCTTGGCCGCCTTGCCCTGC---CGC 58419

Qy     21 GlySerArgGlyPheGlyMetPheTyrAlaValArgArgLysArgLysThrGlyValPhe 40
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  58420 TGCTCTCACGGGTTCGGGATGTTCATGCCGTGAGGAGGCGGCAAGACCGGGGTCTTT 58479

Qy     41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgArgPheProAlaAlaPheLysLys 60
          |||||||||| | : ::|||||||||| |||||||||||||||
Db  58480 CTGACCTGGAATGAGTGCAAAGCACAGTGGACCGGTTCCTGCTGCCAGATTAAGAAG 58539

Qy     61 PheAlaThrGlnAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  58540 TTTGCCACAGAGGATGAGGCCTTTGTCAGGAAATCTGCAAGCCCCGAAGTTTCA 58599

Qy     81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db  58600 GAAGGGCAGGAGAAAATCAACATGGACAAGAATCAGAGACGAAAGCCACAAGCGACTTCGT 58659

Qy    101 GluProLeuAspGlyAspTyrAlaLysGluSerArgAlaGlnProTyrAlaLysHisMetLysPro 120
          |||||||||||  | :::|||||||||||||||||||||||||||
Db  58660 GAGCCACTGGATGGAGATGGAGATGGAAGCCGCAGAGCCATATGC-AAGCACATGAAGCCG 58718

Qy    121 SerValGluProValProProValSerArgArgAspThrPheSerTyrMetGlyAspPheVal 140
          |||||||||||||||||||||||||||||||||||||||||||||||
Db  58719 AGGGTGAAGCCCGTCCCGCCCTCCAGTTAGCAGAGACACGTTTCCTACATGGGAGACTTCGTC 58778

Qy    141 ValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGlyIle 160
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  58779 GTCGTCTCACGCTGGTCTGCTGCTCCCAGTAATGGGCGTAGGAGGGCCACGAGCAGGAATC 58838

Qy    161 GlyValTyrTyrProGlyHisProLeuAsnValGlyIleAlaGlyLeuProGlyArgGln 180
          |||:||||||||||||||||||||||||||||||||||||||||||||
Db  58839 CGCGTTTACTGGGCGCGGGCTATCCTTTAAATGTAGGCATTAGAGACTTCCTGGGTGGCAG 58898

Qy    181 ThrAsnGlnArgAlaGluIleHisAlaAlaAlaCysLysAlaIleGluGluLysThrGln 200
          |||||||||||||||||||||||||||||||||||||||||||||
Db  58899 ACAAACCAAAGAGCCGGAAATTCATGCAGCCTACAAGGCCATTGAACAGCAAAGACTCAA 58958

Qy    201 AsnIleAsnLysLeuValLeuThrTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220
          ||||||||||||||||||||||||||||||||||||||||
Db  58959 AACATCAATAAACTGGTTCTGTATACAGACAGTATGTTTACGATAAATGGGTGT----AAC 59014

Qy    221 TrpValGlnGlyTrpLysLysAsnGlyValTrpLysThrSerArgAlaGlyValIleGluAsn 240
          |||||||||||||||||||||||||||||||||||||||||||||||
Db  59015 TGGGTTCAAGGTTGGAAGAAAAATGGATGGAAGACAAGTGCAGGGCAGGAGAGTGATCAAC 59074

Qy    241 LysGluAspPheValAlaLeuGluThrGlnIleMetAspIleGlnTrp------ 258
          ||||||||||||||||||
Db  59075 AAAGAGGACTTCGTGCACTGGAGAGGCTGACCCAGGGGATGGACATTCAGTGGGCGAGT 59134

Qy    259 --------MetHisValPro 262
               :::|||||||||
Db  59135 ATCTTAAATGTTCATGTCCCA 59155
```

RESULT 14

LOCUS   AC107926   176040 bp   DNA   linear   PRI 01-JUL-2002
DEFINITION Homo sapiens, clone RP11-381P6, complete sequence.
ACCESSION   AC107926
VERSION   AC107926.6 GI:21637512
KEYWORDS   HTG.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 176040)
AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
TITLE   Homo sapiens, clone RP11-381P6
JOURNAL   Unpublished

REFERENCE   2 (bases 1 to 176040)
AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C.,
            Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., Meldrin,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE   Direct Submission
JOURNAL   Submitted (24-JAN-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE   3 (bases 1 to 176040)
AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,
            Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N.,
            Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrin,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C.,
            Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S.,
            Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
            Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE   Direct Submission
JOURNAL   Submitted (15-JUN-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE   4 (bases 1 to 176040)
AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
TITLE   Direct Submission
JOURNAL   Submitted (01-JUL-2002) Whitehead Institute/MIT Center for Genome

COMMENT   Research, 320 Charles Street, Cambridge, MA 02141, USA
           On Jul 1, 2002 this sequence version replaced gi:21427785.
           All repeats were identified using RepeatMasker:
           Smit, A.F.A. & Green, P. (1996-1997)
           http://ftp.genome.washington.edu/RM/RepeatMasker.html
           -------- Genome Center
               Center: Whitehead Institute/ MIT Center for Genome Research
               Center code: WIBR
               Web site: http://www.seq.wi.mit.edu
               Contact: sequence_submissions@genome.wi.mit.edu
           -------- Project Information
               Center project name: L23134
               Center clone name: 381_P_6

           Only the first 176.0 kb of this clone are being submitted.
           The remainder overlaps accession number AC090774 [WICGR project
           L12729].

FEATURES   Location/Qualifiers
     source   1..176040
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="RP11-381P6"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   7..235
                     /rpt_family="AluJo"
     repeat_region   477..662
                     /rpt_family="HAL1"
     repeat_region   1499..1785
                     /rpt_family="AluSc"
     repeat_region   1793..1850
                     /rpt_family="L1MB5"
     repeat_region   1851..2131
                     /rpt_family="AluY"
     repeat_region   2135..2163
                     /rpt_family="AluY"
     repeat_region   2164..2936
                     /rpt_family="(CAA)n"
     repeat_region   2164..2936
                     /rpt_family="L1MB5"
     repeat_region   3425..3575
                     /rpt_family="L1MB5"
     repeat_region   4063..4183
                     /rpt_family="MIR"
     repeat_region   4740..4778
                     /rpt_family="GA-rich"
     repeat_region   4740..4778
                     /rpt_family="AT_rich"
     repeat_region   complement(4785..4940)
                     /rpt_family="FRAM"
     repeat_region   5727..6082
                     /rpt_family="L1MC4a"
     repeat_region   6096..6306
                     /rpt_family="L1MC4a"
     repeat_region   6447..7402
                     /rpt_family="L1MB7"
     repeat_region   7403..7706
                     /rpt_family="AluSg"
     repeat_region   7707..7750
                     /rpt_family="L1MB7"
     repeat_region   7751..8105
                     /rpt_family="AluY"
     repeat_region   8106..8680
                     /rpt_family="L1MB7"
     repeat_region   9419..9524
                     /rpt_family="MER103"
     repeat_region   10074..10358
                     /rpt_family="L1MC5"
     repeat_region   11022..11208
                     /rpt_family="L2"
     repeat_region   11223..11436
                     /rpt_family="MIR"
     repeat_region   12191..12212
                     /rpt_family="(CA)n"
     repeat_region   complement(13126..13321)
                     /rpt_family="MLT1B2"

     repeat_region   13920..14143
                     /rpt_family="L2"
     repeat_region   14198..15784
                     /rpt_family="L1PA16"
     repeat_region   15787..15823
                     /rpt_family="(TG)n"
     repeat_region   complement(16696..16840)
                     /rpt_family="FRAM"
     repeat_region   complement(17030..17339)
                     /rpt_family="AluSg"
     repeat_region   17575..17802
                     /rpt_family="L1ME2"
     repeat_region   17839..17938
                     /rpt_family="L1ME2"
     repeat_region   17935..18247
                     /rpt_family="AluSg"
     repeat_region   18248..18793
                     /rpt_family="L1ME2"
     repeat_region   19076..19398
                     /rpt_family="LTR35"
     repeat_region   19399..19633
                     /rpt_family="L1PA13"
     repeat_region   20008..20393
                     /rpt_family="AluY"
     repeat_region   20402..20535
                     /rpt_family="L1MC/D"
     repeat_region   21200..21338
                     /rpt_family="LTR67"
     repeat_region   21344..21474
                     /rpt_family="L1MC/D"
     repeat_region   complement(21546..21680)
                     /rpt_family="L2"
     repeat_region   complement(22297..22593)
                     /rpt_family="AluSq"
     repeat_region   25626..25660
                     /rpt_family="(TG)n"
     repeat_region   25875..25934
                     /rpt_family="AT_rich"
     repeat_region   complement(26234..26533)
                     /rpt_family="AluY"
     repeat_region   26951..27234

Alignment Scores:
Pred. No.:
Score: 1.26e-71   Length: 176040
Percent Similarity: 1145.50   Matches: 233
Best Local Similarity: 89.51%   Conservative: 6
Query Match: 87.27%   Mismatches: 21
DB: 74.09%   Indels: 10
                                9   Gaps: 3

US-10-054-313-1 (1-286) x AC107926 (1-176040)

Qy   1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaAlaLeuProCysArgArg   20
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   77485 ATGAGCCGGCTTCTGTTCCTGGCCGC-AGAGTCGGCCTTGGCCCGCCTTCCCTGC---CGC   77540

Qy   21 GlySerArgGlyPheGlyMetPheTyrAlaValArgArgGlyArgLysThrGlyValPhe   40
             ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   77541 TGCTCTCACGGGTTCGGATGTTCTATGCCGTGAGGAGGGGCGGCAAGACCGGGGTCTTT   77600

Qy   41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAArgPheLysLys   60
             |||||||||||::::||||||||||||||||||||||||||||||||||||||
Db   77601 CTGACCCGGAATGAGTGCAAAGCACAGGTGGACCGGTTTCCTGCTGCCAGATTTAAGAAG   77660

Qy   61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer   80
             |||||||||||||||||||||||||||||||||||||
Db   77661 TTTGCCACAGAGGATGAGACCTGGGACTTTGTCAGGAAATCTGCAAGCCCGGAAGTTTCA   77720

Qy   81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg   100
             ||||||||||||||| || |||||||
Db   77721 GAAGGGCACGAAAATCAACATGGACAAGATCAGAGACCAAAGCCAGCAAGCGACTCCGT   77780

Qy   101 GluProLeuAspGlyAspGlyHisSerAlaGlnProTyrAlaLysHisMetLysPro   120

```
Contact: humquery@sanger.ac.uk
--------------
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest, except on the rare
occasion of the clone being a YAC.
RP4-667H12 is from the library RPCI-4 constructed by the group of
Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pCYPAC2.
```

FEATURES   Location/Qualifiers
source   1..131239
    /organism="Homo sapiens"
    /mol_type="genomic DNA"
    /db_xref="taxon:9606"
    /chromosome="1"
    /map="q32.1-41"
    /clone="RP4-667H12"
    /clone_lib="RPCI-4"

BASE COUNT   35649 a   28535 c   29022 g   38033 t
ORIGIN

Alignment Scores:
Pred. No.:   131239
Score: 1016.00
Percent Similarity: 79.37%
Best Local Similarity: 73.08%
Query Match: 65.72%

Length: 131239
Matches: 209
Conservative: 18
Mismatches: 58
Indels: 2
Gaps: 1

US-10-054-313-1 (1-286) x HS667H12 (1-131239)

```
Qy      1 MetSerTrpLeuPheLeuAlaHisArgValAlaAlaLeuAlaAlaLeuProCysArgArg  20
            ||||||    |||||
Db  85425 ATGTGCCAGCTTCTGTTCCTGCCCACAGAGTCACCTTGATAGCTGTGCGCTGCCACAGG 85484

Qy     21 GlySerArgGlyPheGlyMetPheTyrAlaValValArgGlyAlaArgThrGlyValValPhe  40
            ||||||||||||||||||||    |||||
Db  85485 GACTCTCGCAGCCTCCAGGATGGTCTATGCTGTGGGGAGGGCCTGCAGGACTGGGGTCTTT 85544

Qy     41 LeuThrTrpAsnGluCysArgValAlaGlnValAspArgPheProAlaAlaArgPheLysLys  60
            |||:||    ||||||||||    ||||    ||||
Db  85545 CTGACCTGGAATGAATATAGAGCACAGGTGGACCATTCCTGCTGCCAGATTTAAGAAG 85604

Qy     61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer  80
            ||||:||||||    ||||||    ||||||
Db  85605 TTTGCCACAGAAAATGAGGCCTGGGCTTTCTCAGGAAATCTGCAGCCCAGAAGGTTCA 85664

Qy     81 GluGlyHisAsnGlnHisGlyHisGlyGluSerGluAlaAlaLysProGlyLysArgLeuArg  100
            ||||||    ||||::||||||    |||||
Db  85665 GAAGGGCAGAAAAATAAACATGTACAAGAATCACAAATGAGAAGCCAGCAAGGGACTCTGT 85724

Qy    101 GluProLeuAspGlyAspTyrAlaAspGluHisGluSerAlaGlnProTyrAlaLysMetLysPro  120
            ||||:||    ||||    ||||||:||
Db  85725 GAGCCACTGAATGGAGCATGGAGATGAAAGTCAGAGTCTGTGCAAAGCCTGTGAAGCAA 85784

Qy    121 SerValGluProAlaProProValSerArgArgAspThrPheSerTyrMetGlyAspPheVal  140
            ::|||    ||||||    |||||:|||
Db  85785 ATCATAGAGCCCGGCACCTCAGTGAGCAGAGACA--TTTTCTTACATCGGAGGCTTTGTC 85842

Qy    141 ValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProAlaAlaGlyIle  160
            ||||||||    ||||||    :|||
Db  85843 GTCGTCCACACTGATGGCTACTGCTCCAGTAATGGTGGAGGAGCACTAGTGGGAACT 85902

Qy    161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyIleArgGlyLeuProGlyArgGln  180
            |||||||||||||||||||||||||||||||
Db  85903 GACGTTTACTGGGGCCAGGCCATCCTTAAATGTAGGCATTAGACTTCCTGGGTGACAG 85962

Qy    181 ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaIleGluGlnAlaLysThrGln  200
```

(left column)

```
Db  77781 GAGCCACTGGATGGAGATGGAGAAGCCGAGAGCCATATGC-AAGCACATGAAGCCG 77839
            |||||||    ||||||||||    :|||||
Qy    121 SerValGluProAlaProProValSerArgArgAspThrPheSerTyrMetGlyAspPheVal  140

Db  77840 AGGGTGAAGCCGGCGCCCAGTTAGCACGACACAGTTTCCTACATGGGAGACTTCGTC 77899
            ||:|||    |||||||||||||||||||||
Qy    141 ValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProAlaAlaGlyIle  160

Db  77900 GTCGTCCACGCTGATGGCTGTCGTCTCCAGTAATGGCGTAGAAGGCCACGACCAGAAATC 77959
            |||||||:::||||:||    |||||
Qy    161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyIleArgGlyLeuProGlyArgGln  180

Db  77960 CGCGTTTACTGGGGCCGGGCTATCCTTAAATGTAGGCATTAGACTTCCTGGGTGGCAG 78019
            ||||||||||||||||    |||||
Qy    181 ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaIleGluGlnAlaLysThrGln  200

Db  78020 ACAAACCAAAGAGCGGAAATTCATGCAGCCTACAAGCCATTGAACAAGCAAAGACTCAA 78079
            |||||||||||||||||||    |||||
Qy    201 AsnIleAsnLysLeuValTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn  220

Db  78080 AACATCAATAAACTGGTTCTGTATACAGACAGTATGTTTACGATAAATGGTGT----AAC 78135
            |||||||:::||||    |||||
Qy    221 TrpValGlnGlyTrpLysAsnGlyThrSerSerAlaGlyValLysGluValIleAsn  240

Db  78136 TGGGTTCAAGGTTGGAAGAAAATGATGGAGACAAGTGCAGGGAAAGAGGTGACCAAC 78195
            |||||||||||||||||    |||||
Qy    241 LysGluAspPhePheValAlaLeuGluThrGlnGlyLeuThrGlnGlyMetAspIleGlnTrp------  258

Db  78196 AAAGAGGACTTCTGTGGCCACTGGAGAGGCTGACCCACCCAGGGATGGACATTCAGTGGGCGAGT 78255
            ||:||||||||    :|||
Qy    259 --------MetHisValPro  262

Db  78256 ATCTTAAATGTTCATGTCCCA 78276
```

```
RESULT 15
HS667H12
LOCUS      HS667H12       131239 bp   DNA     linear   PRI 04-MAR-2003
DEFINITION Human DNA sequence from clone RP4-667H12 on chromosome 1q32.1-41,
           complete sequence.
ACCESSION  AL035414
VERSION    AL035414.30  GI:6822166
KEYWORDS   HTG.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1  (bases 1 to 131239)
AUTHORS    Ellington.A.
TITLE      Direct Submission
JOURNAL    Submitted (04-MAR-2003) Wellcome Trust Sanger Institute, Hinxton,
           Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
           humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT    On Jan 30, 2000 this sequence version replaced gi:6782338.
           During sequence assembly data is compared from overlapping clones.
           Where differences are found these are annotated as variations
           together with a note of the overlapping clone name. Note that the
           variation annotation may not be found in the sequence submission
           corresponding to the overlapping clone, as we submit sequences with
           only a small overlap as described above.
           The following abbreviations are used to associate primary accession
           numbers given in the feature table with their source databases:
           Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
           on the WORMPEP database can be found at
           http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
           was generated from part of bacterial clone contigs of human
           chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
           Group. Further information can be found at
           http://www.sanger.ac.uk/HGP/Chr1
           --------- Genome Center
           Center: Wellcome Trust Sanger Institute
           Center code: SC
           Web site: http://www.sanger.ac.uk
```

```
Db  85963 ACAAACCAAAGAGCAGAAAATCATGCAGCCTGCAACACCATTGAACAAGCAAAGGCTGAA  86022
          |||||||||||||||     ||||||||||                    |||||||| : ::

Qy    201 AsnIleAsnLysGluValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn  220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  86023 AACATCAATAAATTGGTTCTGTACACAGATAGTATGTTTGCTATAAATGGTATAACTAAC  86082

Qy    221 TrpValGlnGlyTrpLysLysSerAlaGlyTrpLysThrSerAlaGlyLysGluValIleAsn  240
          |||||||||||||||     ||||||| |||:::||
Db  86083 TGGGTTCAAGGTAGGAAGAAGAACAGGTGGAAGACAAGTACAGGGAAATAGGTGATGAAC  86142

Qy    241 LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHis  260
          ||:::||||||||:||||||||||||||||| |||||||||||||:||:
Db  86143 AAAAAGGACTTTGTGGCACTGGAGAAGCTCAGTCAAGGCCATGGACATTCAGTGGATACTT  86202

Qy    261 ValProGlyHisSerGlyIlePheIleGlyAsnGluGluAlaAspArgLeuAlaArgGluGly  280
          |||||||||||||||       |||||||||||||||||||||||||||||    |||
Db  86203 GTTCCTGGTCATCCAGGGTGTATTGGCAATGAAGAAGCTGATAGATTAGCAAGA--GGA  86259

Qy    281 AlaLysGlnSerGluAsp  286
          ||||||||||||||||||
Db  86260 GCTAAACATTCTGAAGAC  86277
```

Search completed: December 17, 2003, 20:35:04
Job time : 4343 secs

OM protein - nucleic search, using frame_plus_p2n model

```
             GenCore version 5.1.6
  Copyright (c) 1993 - 2003 Compugen Ltd.

Run on:    December 17, 2003, 18:02:28 ; Search time 318 Seconds
                                         (without alignments)
                                         2427.796 Million cell updates/sec

Title:         US-10-054-313-1
Perfect score: 1546
Sequence:      1 MSWLLFLAHRVALAALPCRR........FIGNEEADRLAREGAKQSED 286

Scoring table: BLOSUM62
               Xgapop 10.0 , Xgapext 0.5
               Ygapop 10.0 , Ygapext 0.5
               Fgapop  6.0 , Fgapext 7.0
               Delop   6.0 , Delext  7.0

Searched:      2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters:     5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=x1h
-Q=/cgn2_1/USPTO_spool/US10054313/runat_17122003_150742_24169/app_query.fasta_1.455
-DB=N_Geneseq_19Jun03 -QFMT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USR=US10054313_@GN_1_312 @runat_17122003_150742_24169 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :   N_Geneseq_19Jun03:*
     1: /SIDS1/gcgdata/geneseq/geneseq-embl/NA1980.DAT:*
     2: /SIDS1/gcgdata/geneseq/geneseq-embl/NA1981.DAT:*
     3: /SIDS1/gcgdata/geneseq/geneseq-embl/NA1982.DAT:*
     4: /SIDS1/gcgdata/geneseq/geneseq-embl/NA1983.DAT:*
     5: /SIDS1/gcgdata/geneseq/geneseq-embl/NA1984.DAT:*
     6: /SIDS1/gcgdata/geneseq/geneseq-embl/NA1985.DAT:*
     7: /SIDS1/gcgdata/geneseq/geneseq-embl/NA1986.DAT:*
     8: /SIDS1/gcgdata/geneseq/geneseq-embl/NA1987.DAT:*
     9: /SIDS1/gcgdata/geneseq/geneseq-embl/NA1988.DAT:*
    10: /SIDS1/gcgdata/geneseq/geneseq-embl/NA1989.DAT:*
    11: /SIDS1/gcgdata/geneseq/geneseq-embl/NA1990.DAT:*
    12: /SIDS1/gcgdata/geneseq/geneseq-embl/NA1991.DAT:*
    13: /SIDS1/gcgdata/geneseq/geneseq-embl/NA1992.DAT:*
    14: /SIDS1/gcgdata/geneseq/geneseq-embl/NA1993.DAT:*
    15: /SIDS1/gcgdata/geneseq/geneseq-embl/NA1994.DAT:*
    16: /SIDS1/gcgdata/geneseq/geneseq-embl/NA1995.DAT:*
    17: /SIDS1/gcgdata/geneseq/geneseq-embl/NA1996.DAT:*
    18: /SIDS1/gcgdata/geneseq/geneseq-embl/NA1997.DAT:*
    19: /SIDS1/gcgdata/geneseq/geneseq-embl/NA1998.DAT:*
    20: /SIDS1/gcgdata/geneseq/geneseq-embl/NA1999.DAT:*
    21: /SIDS1/gcgdata/geneseq/geneseq-embl/NA2000.DAT:*
    22: /SIDS1/gcgdata/geneseq/geneseq-embl/NA2001A.DAT:*
    23: /SIDS1/gcgdata/geneseq/geneseq-embl/NA2001B.DAT:*
    24: /SIDS1/gcgdata/geneseq/geneseq-embl/NA2002.DAT:*
    25: /SIDS1/gcgdata/geneseq/geneseq-embl/NA2003.DAT:*
```

and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Query Length	DB	ID	Description
	1	1526	98.7	2129	23	ABV30263	Human prostate exp
	2	1523	98.5	11150	21	AAZ51265	Human RNA-associat
	3	905	58.5	545	23	ABV45787	Human prostate exp
	4	666.5	43.1	691	23	ABV15988	Human colon cancer
	5	508.5	32.9	389	21	AAH30241	Bovine EST associa
	6	434	28.1	473	25	ABX41815	Drosophila melanog
	7	422	27.3	1071	23	ABL03247	Bovine EST associa
	8	399	25.8	407	25	ABX38735	Frog embryonic gen
	9	364	23.5	764	24	ABS77198	Drosophila melanog
	10	356.5	23.1	3290	23	ABL03246	Genomic fragment #
c	11	272	17.6	119211	22	AAF28853	Human gene signatu
	12	244	15.8	378	16	AAT24248	Polynucleotide seq
c	13	239	15.5	7874	20	AAX20568	Haemophilus influe
	14	236.5	15.3	462	24	AAF94380	DNA encoding Haemo
	15	236.5	15.3	462	22	ABK64944	Haemophilus influe
	16	236.5	15.3	1001	22	AAF91449	Haemophilus influe
	17	236.5	15.3	1001	24	ABK37829	DNA sequence upstr
	18	236.5	15.3	1830121	17	AAT42063	Haemophilus influe
c	19	235.5	15.2	2048	20	AAX98244	Contig 1344 identi
	20	234.5	15.2	435	25	ABZ40958	N. gonorrhoeae nuc
	21	229.5	14.8	36471	21	AAA81453	N. meningitidis pa
	22	229.5	14.8	349980	21	AAF21611	Neisseria meningit
	23	229.5	14.8	837096	21	AAA81489	N. meningitidis pa
	24	221.5	14.3	801	20	AAX98246	Nucleotide sequene
	25	213.5	13.8	465	16	AAQ96258	E.coli ribonucleas
	26	211.5	13.7	465	16	AAQ96267	E.coli ribonucleas
	27	209.5	13.6	757	12	AAQ12780	Encodes stabilised
c	28	203	13.1	349980	24	ABQ81849	Bifidobacterium lo
	29	187	13.1	32502	23	AAS59542	Propionibacterium
	30	182.5	11.8	6397	23	ABL03444	Drosophila melanog
	31	176.5	11.4	560	13	AAQ20910	Thermophilic bacte
	32	174	11.3	501	16	AAT00881	Thermus flavus rec
	33	174	11.3	501	17	AAT17719	Thermus flavus AT-
	34	174	11.3	552	16	AAT00882	Thermus thermophil
	35	174	11.3	552	17	AAT17720	Thermus thermophil
	36	162.5	10.5	644	21	AAF12943	Aspergillus oryzae
	37	160	10.3	8088	21	AAZ45540	Complete nucleotid
	38	160	10.3	8535	15	AAQ73731	GaLV SEATO genome.
	39	157	10.2	8060	18	AAT74811	Porcine retrovirus
	40	157	10.2	8060	22	AAF77725	Tsukuba-1 cDNA. P
c	41	156	10.1	8202	17	AAT09280	Novel AMP/MCF viru
	42	155	10.0	8440	17	AAT13265	FeLV F6A provirus
	43	154	10.0	3612	25	ABS57820	MoMLV retrovirus p
	44	154	10.0	3643	21	AAA87841	MLV reverse transc
	45	154	10.0	5178	21	AAA15660	Gibbon leukaemia v

ALIGNMENTS

```
RESULT 1
ABV30263
ID   ABV30263 standard; cDNA; 2129 BP.
XX
AC   ABV30263;
XX
DT   16-SEP-2002  (first entry)
XX
DE   Human prostate expression marker cDNA 30254.
XX
KW   Human; prostate cancer; cytostatic; carcinogen; pharmacodyanamic marker;
KW   pharmacogenomic marker; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200160860-A2.
XX
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,

```
PD   23-AUG-2001.
XX
PF   20-FEB-2001; 2001WO-US05171.
XX
PR   17-FEB-2000; 2000US-183319P.
PR   16-MAR-2000; 2000US-189862P.
PR   25-MAY-2000; 2000US-207454P.
PR   09-JUN-2000; 2000US-211314P.
PR   18-JUL-2000; 2000US-219007P.
PR   13-DEC-2000; 2000US-255281P.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Schlegel R, Endege WO, Monahan JE;
XX
DR   WPI; 2001-662795/76.
XX
PT   Novel isolated nucleic acid molecule associated with cancerous state of
PT   prostate cells and correlating with presence of prostate cancer, useful
PT   for detecting presence of prostate cancer, stage of prostate cancer -
XX
PS   Claim 1; Page 6567-6568; 11750pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule (I) comprising
CC   a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC   specification or its complement. (I) is useful for:
CC   (a) assessing whether a patient is afflicted with prostate cancer;
CC   (b) monitoring the progression of prostate cancer in a patient;
CC   (c) assessing the efficacy of a test compound to inhibit prostate
CC       cancer in a patient;
CC   (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC       in a patient;
CC   (e) selecting a composition for inhibiting prostate cancer in a patient;
CC   (f) assessing the prostate cell carcinogenic potential of a compound;
CC   (g) determining whether prostate cancer has metastasized in a patient;
CC   (h) assessing the aggressiveness or indolence of prostate cancer in a
CC       patient;
CC   (I) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX
SQ   Sequence 2129 BP; 609 A; 510 C; 578 G; 427 T; 5 other;

Alignment Scores:
Pred. No.:
Score:                  5.9e-126     Length:        2129
Percent Similarity:     1526.00      Matches:       282
Best Local Similarity:  99.30%       Conservative:  2
Query Match:            98.60%       Mismatches:    2
DB:                     98.71%       Indels:        0
                         23          Gaps:          0

US-10-054-313-1 (1-286) x ABV30263 (1-2129)

Qy    1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaAlaAlaLeuProCysArgArg   20
Db  112 ATGAGCTGGCTTCTTCGTTCCTGGCCCACAGAGTCGCCTTGGCCGCCGCCTTGCCCTGCCGCGGC 171

Qy   21 GlySerArgGlyPheMetPheTyrAlaValArgArgValArgLysThrGlyValPhe         40
Db  172 GGCTCCTCGCCGGGTTCGGGATGTTCTATGCCGTGAGGAGGGTGAGGAAGACCGGGGTCTTT   231

Qy   41 LeuThrTrpAsnGluCysArgArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys   60
Db  232 CTGACCTGGAATGAGTGCAGAGCACAGGTGGACCGGTTTCCTGCTGCCAGATTTAAGAAG     291

Qy   61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer      80
Db  292 TTTGCCACAGAGGATGAGGGCTGGGCCTTTGTCAGGAAATCTGCAAGCCCGGAAGTTTCA     351

Qy   81 GlyHisGlyAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg        100
Db  352 GAAGGGCCATGAAAATCAACATGGACAAGAATCGGAGGCAGAAGCCAGCCAGCGACTCCGT    411

Qy  101 GluProLeuAspGlyAspTyrArgAlaGlnProTyrArgAlaLysHisMetLysPro       120

Db  412 GAGCCACTGGATGGAGATGGACATGAAAGGCGCAGAGCCGTATGCAAAGCACATGAAGCCG   471
Qy  121 SerValGluProAlaProProValSerArgArgAspThrPheSerTyrMetGlyAspPheVal  140
Db  472 AGCGTGGAGCCGGCCGCCTCCAGTTAGCAGAGACACGTTTTCCTACATGGGAGACTTCGTC   531
Qy  141 ValValTyrThrAspGlyCysGlySerSerAsnGlyArgArgLysProAlaGlyIle        160
Db  532 GTCGTCTACACTGATGGCTGCTGCTCCAGTAATGGGCGTAGAAGGCCGCGAGCAGGAAATC   591
Qy  161 GlyValTyrArgGlyProGlyHisProLeuAsnValGlyValIleArgLeuProGlyArgGln  180
Db  592 GGCGTTTACTGGGGCCAGGCATCCTTAAATGTAGGCATTAGGACTTCCTGGGCGCAG       651
Qy  181 ThrAsnGlnArgAlaAlaGluIleHisAlaAlaCysLysAlaIleGluGlnAlaLysThrGln  200
Db  652 ACAAACCAAGAGCCGAAATCATGCAGCCTGCAAAGCAATTGACAAGCAAAGACTCAA       711
Qy  201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn     220
Db  712 AACATCAATAAACTGGTTCTGTATACAGACAGTATGTTTACGATAAATGGTATAACTAAC    771
Qy  221 TrpValGlnGlyValTrpLysLysAsnGlyTrpLysThrSerAlaGlyIleLysGluValIleAsn 240
Db  772 TGGGTTCAAGGTGGAAGAAAAATGGGTGGAAGACAAGTGCAGGGAAAGAGGTGATCAAC     831
Qy  241 LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpGlnMetHis  260
Db  832 AAAGAGGACTTTGTGGCACTGGAGAGGCTTACCCAGGGGATGGACATTCAGTGGGATGCAT   891
Qy  261 ValProGlyHisSerGlyPheIleGlyAsnGluAlaAspArgLeuAlaArgGluGly        280
Db  892 GTTCCTGGTCATTCGGGATTTATAGGCAATGAAGAAGAGCTGACAGATTAGCCAGAGAGGA   951
Qy  281 AlaLysGlnSerGluAsp                                              286
Db  952 GCTAAACAATCGGAAGAC                                              969

RESULT 2
AAZ51265
ID   AAZ51265 standard; cDNA; 1150 BP.
XX
AC   AAZ51265;
XX
DT   06-JUN-2000  (first entry)
XX
DE   Human RNA-associated protein-16 (RNAAP-16) encoding cDNA.
XX
KW   RNA-associated protein; RNAAP; human; clone 2073417; cytostatic;
KW   immunosuppressive; antiinflammatory; antiallergic; neuroprotective;
KW   antiarteriosclerotic; hepatotropic; antipsoriatic; virucide; anti-HIV;
KW   antiallergic; antirheumatic; antiarthritic; opthalmological; autoimmune;
KW   antimicrobial; cell proliferative disorder; inflammation; cirrhosis;
KW   actinic keratosis; bursitis; arteriosclerosis; artherosclerosis;
KW   hepatitis; myelofibrosis; primary thrombocythemia; psoriasis; cancer;
KW   mixed connective tissue disease; MCTD; HIV; uveitis; Crohn's disease;
KW   allergy; rheumatoid arthritis; parasitic infection; ss.
XX
OS   Homo sapiens.
XX
FH   Key              Location/Qualifiers
FT   CDS              85..945
FT                    /tag=  a
FT                    /product= "Human RNA-associated protein-16"
FT                    /note= "Derived from ISLTNOT01 library"
FT   sig_peptide      85..138
FT                    /tag=  b
FT   mat_peptide      139..942
FT                    /tag=  c
FT                    /product= "Mature human RNA-associated protein-16"
XX
PN   WO20011171-A2.
XX
```

PD   02-MAR-2000.
XX
PF   20-AUG-1999;   99WO-US19361.
XX
PR   21-AUG-1998;   98US-0097550.
PR   12-JAN-1999;   99US-0115639.
XX
PA   (INCY-) INCYTE PHARM INC.
XX
PI   Hillman JL, Yue H, Tang YT, Corley NC, Guegler KJ, Gorgone GA;
PI   Patterson C, Baughn MR, Lal P, Bandman O, Reddy R, Azimzai Y;
PI   Shih LL, Yang J, Lu DAM;
XX
DR   WPI; 2000-237651/20.
DR   P-PSDB; AAY70235.
XX
PT   Human RNA-associated proteins useful in diagnosing, treating and
PT   preventing cell proliferative, autoimmune, inflammatory and infectious
PT   disorders -
XX
PS   Claim 9; Page 117-118; 123pp; English.
XX
CC   The present sequence is the cDNA encoding human RNA-associated protein-16
CC   (RNAAP-16), identified in Incyte clone 2073417, derived from ISLTNOT01
CC   library. It is expressed in nervous, reproductive, gastrointestinal,
CC   cardiovascular and haematopoietic/immune tissues. It has cytostatic,
CC   immunosuppressive, antiinflammatory, antiarteriosclerotic, hepatotropic,
CC   keratolytic, neuroprotective, antipsoriatic, anti-HIV, antiallergic,
CC   antirheumatic, virucide, antiarthritic, opthalmological and antimicrobial
CC   activity. RNAAP antibodies are useful for diagnosis of diseases
CC   associated with altered expression or activity of RNAAP. It is used to
CC   treat cell proliferative, autoimmune, inflammatory and infectious
CC   disorders, like actinic keratosis, bursitis, arteriosclerosis,
CC   artherosclerosis, cirrhosis, hepatitis, myelofibrosis, mixed connective
CC   tissue disease (MCTD), psoriasis, primary thrombocythemia and cancer,
CC   HIV, allergies, rheumatoid arthritis, uveitis, Crohn's disease, and
CC   bacterial, viral and parasitic infections.
XX
SQ   Sequence 1150 BP; 302 A; 250 C; 352 G; 246 T; 0 other;

Alignment Scores:
Pred. No.:
Score: 5.1e-126   Length: 1150
Percent Similarity: 99.30%   Matches: 281
Best Local Similarity: 98.25%   Conservative: 3
Query Match: 21   Mismatches: 2
DB: Indels: 0
                                          Gaps: 0

US-10-054-313-1 (1-286) x AAZ51265 (1-1150)

Qy   1 MetSerTrpLeuPheLeuAlaHisArgValAlaLeuAlaAlaLeuProCysArgArg 20
Db   85 ATGAGCTGGCTTCTGTTCCTGGCCCACAGAGTCGCCCTTGCCCTGCCGCCGC 144

Qy   21 GlySerArgGlyPheGlyMetPheTyrAlaValArgArgGlyValArgLysThrGlyValPhe 40
Db 145 GGCTCTCGCGGGTTCGGGATGTTCTATGCCGTGAGGAGGGGCGTCAAGACCGGGGTCTTT 204

Qy   41 LeuThrTrpGluGluCysArgAlaGlnValAlaAspArgPheProAlaAlaArgPheLysLys 60
Db 205 CTGACCTGGGAAGAGTGCAGGGCGCAGGTGGCAGACCGGTTCCTGCTGCCAGATTTAAGAAG 264

Qy   61 PheAlaThrGluGluAspAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80
Db 265 TTTGCCACAGAGGATGAGGCCTGGGCCTTTGTCAGGAAATCTGCAAGCCCGGAAGTTTCA 324

Qy   81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100
Db 325 GAAGGGCATGAAAATCAACATGGAGAAAAGCCGAAAGCCAGCAAGCGACTTCCGT 384

Qy 101 GluProLeuAspGlyAspGlnHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120
Db 385 GAGCCACTGGATGGAGATGGACAACATGAAAGCCCAGCCGTATGCAAAGCACATGAAGCCG 444

Qy 121 SerValGluProAlaProProValSerArgArgAspThrPheSerTyrMetGlyAspPheVal 140
Db 445 AGCATGGAGCCGGCGCCTCCAGTTAGCAGAGACACGTTTTCCTACATGGGAGACTTCGTC 504

Qy 141 ValValTyrThrAspGlyTyrCysCysSerCysAsnGlyArgArgLysProArgAlaGlyIle 160
Db 505 GTCGTTCACACTGATGGCTGCTGCTGCTCCAGTAATGGCGTAGAAGGCCGCAGCAGGAATC 564

Qy 161 GlyValTyrTrpThrGlyProGlyHisProLeuAsnValGlyTyrArgGluProGlyArgGln 180
Db 565 GGCGGTTTACTGGGGACCAGGCCATCCTTTAAATGTAGGCATTAGACTTCCTGGGCGGCAG 624

Qy 181 ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaIleGluGlnGlnAlaLysThrGln 200
Db 625 ACAAACCAAAGAGCGGAAATTCATGCAGCCTTGCAAAGCCATTGAACAAGCAAAGACTCAA 684

Qy 201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220
Db 685 AACATCAATAAACTGGTTCTGTATACAGACAGTATGTTTACGATAAATGGTATAACTAAC 744

Qy 221 TrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysValValIleAsn 240
Db 745 TGGGTTCAAGGTTGGAAGAAAAATGGGTGGAAGACAAGTGCAGGGAAAGAGGTGATCAAC 804

Qy 241 LysGluAspPheValAlaLeuGluArgLeuThrArgGlnGlyMetAspIleGlnTrpMetHis 260
Db 805 AAAGAGGACTTTGTGGCACTGGAGAGGCTTACCCGAGGGATGGACATTCAGTGGATGCAT 864

Qy 261 ValProGlyHisSerGlyPheLeuGlyLysAsnGluGlnAlaAspArgLeuAlaAargGluGly 280
Db 865 GTTCCTGGTCATTCGGGATTTATAGGCAATGAAGAAGCTGACAGATTAGCCAGAGAAGGA 924

Qy 281 AlaLysGlnSerGluAsp 286
Db 925 GCTAAACAATCGGAAGAC 942

RESULT 3
ABV45787
ID   ABV45787 standard; cDNA; 545 BP.
XX
AC   ABV45787;
XX
DT   16-SEP-2002 (first entry)
XX
DE   Human prostate expression marker cDNA 45778.
XX
KW   Human; prostate cancer; cytostatic; carcinogen; pharmacodyanamic marker;
KW   pharmacogenomic marker; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO20160860-A2.
XX
PD   23-AUG-2001.
XX
PF   20-FEB-2001;   2001WO-US05171.
XX
PR   17-FEB-2000;   2000US-183319P.
PR   16-MAR-2000;   2000US-189862P.
PR   25-MAY-2000;   2000US-207454P.
PR   09-JUN-2000;   2000US-211314P.
PR   18-JUL-2000;   2000US-219007P.
PR   13-DEC-2000;   2000US-255281P.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Schlegel R, Endege WO, Monahan JE;
XX
DR   WPI; 2001-662795/76.
XX
PT   Novel isolated nucleic acid molecule associated with cancerous state of
PT   prostate cells and correlating with presence of prostate cancer, useful

```
PT        for detecting presence of prostate cancer, stage of prostate cancer  -
XX
PS        Claim 1; Page 9051; 11750pp; English.
XX
CC        The invention relates to an isolated nucleic acid molecule (I) comprising
CC        a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC        specification or its complement. (I) is useful for:
CC        (a) assessing whether a patient is afflicted with prostate cancer;
CC        (b) monitoring the progression of prostate cancer in a patient;
CC        (c) assessing the efficacy of a test compound to inhibit prostate
CC        cancer in a patient;
CC        (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC        in a patient;
CC        (e) selecting a composition for inhibiting prostate cancer in a patient;
CC        (f) assessing the prostate cell carcinogenic potential of a compound;
CC        (g) determining whether prostate cancer has metastasized in a patient;
CC        (h) assessing the aggressiveness or indolence of prostate cancer in a
CC        patient;
CC        (I) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX
SQ        Sequence 545 BP; 136 A; 129 C; 178 G; 102 T; 0 other;

Alignment Scores:
Pred. No.:              1.62e-71        Length:              545
Score:                  905.00          Matches:             167
Percent Similarity:     94.44%          Conservative:        3
Best Local Similarity:  92.78%          Mismatches:          10
Query Match:            58.54%          Indels:              0
DB:                     23              Gaps:                0

US-10-054-313-1 (1-286) x ABV45787 (1-545)

Qy     13  LeuAlaAlaLeuProCysArgArgGlySerArgGlyPheGlyMetPheTyrAlaValArg   32
           ::: |||  ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4  ATTGGAGTCCACCGCGGTGGCGCCGAGGTGCGGGTTCGGGATGTTCTATGCCGTGAGG   63

Qy     33  ArgGlyArgGlyThrGlyValPheLeuThrThrAsnGluCysArgArgAlaGlnValAspArg   52
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     64  AGGGGCCGCCAAGACTGGGTCTTTCTGACCTGGAATGAGTGCAGAGCACAGGTGGACCGG   123

Qy     53  PheProAlaAlaArgPheLysPheAlaThrGluAspGluAlaTrpGluAlaPheValArg   72
           |||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    124  TTTCCTGCTGCCAGATTTAAGAAGTTGCCACAGAGGATGAGGCCTGGCCTTTGTCAGG   183

Qy     73  LysSerAlaSerProGluValSerGluGlyHisGluGlnHisGlyGlnGluSerGlu   92
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    184  AAATCTGCAAGCCCCGGAAGTTTCAGAAGGGCATGAAGCATGGACAAGAATCGGAG   243

Qy     93  AlaLysProGlyLysArgLeuArgGluArgGluProLeuAspGlyAspGlyHisGluSerAlaGln   112
           |||||||||||||||||||||||||||||||||||||||||||||||||::::|||
Db    244  GCGAAAGCCAGCAAGCGACTCCGTGAGCCACTGGATGGAGATGACATGAAAGGCAGAG   303

Qy    113  ProTyrAlaLysHisMetLysProSerValGluProAlaProProValSerArgArgAspThr   132
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    304  CCGTATGCAAAGCACATGAAGCCGAGCGTGGAGCCCGCCTCCAGTTAGCAGAGACACG   363

Qy    133  PheSerTyrMetGlyAspPheValValValTyrThrAspGlyCysCysSerSerAsnGly   152
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    364  TTTTCCTACATGGGAGACTTCGTCGTCGTCTACACTGATGGCTGCTGCTCCAGTAATGGG   423

Qy    153  ArgArgLysProArgAlaGlyIleGlyValTyrTrpGlyProGlyHisProLeuAsnVal   172
           |||||||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db    424  CGTAGAAGCCGGCCGCGAGCAGGCAGGAATCGGCGTTACTGGGGCCCGGGCCATCCTTTAAATGTA   483

Qy    173  GlyIleArgLeuProGlyGlyArgGlnThrArgAsnGlnArgArgAlaGlyIleHisAlaAlaCysLys   192
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    484  GGCATTAGACTTCCTGGGCCGGCAGCAAACCAAAGAGCGAAATTCATGCAACCTGCAAA   543

RESULT 4
ABV15988
ID     ABV15988 standard; cDNA; 691 BP.
XX
AC     ABV15988;
```

```
PT        13-SEP-2002 (first entry)
DT
XX
DB        Human prostate expression marker cDNA 15979.
XX
KW        Human; prostate cancer; cytostatic; carcinogen; pharmacodyanamic marker;
KW        pharmacogenomic marker; gene; ss.
XX
OS        Homo sapiens.
XX
PN        WO20016O860-A2.
XX
PD        23-AUG-2001.
XX
PF        20-FEB-2001; 2001WO-US05171.
XX
PR        17-FEB-2000; 2000US-183319P.
PR        16-MAR-2000; 2000US-189862P.
PR        25-MAY-2000; 2000US-207454P.
PR        09-JUN-2000; 2000US-211314P.
PR        18-JUL-2000; 2000US-219007P.
PR        13-DEC-2000; 2000US-25281P.
XX
PA        (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI        Schlegel R, Endege WO, Monahan JE;
XX
DR        WPI; 2001-662795/76.
XX
PT        Novel isolated nucleic acid molecule associated with cancerous state of
PT        prostate cells and correlating with presence of prostate cancer, useful
PT        for detecting presence of prostate cancer, stage of prostate cancer  -
XX
PS        Claim 1; Page 2678; 11750pp; English.
XX
CC        The invention relates to an isolated nucleic acid molecule (I) comprising
CC        a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC        specification or its complement. (I) is useful for:
CC        (a) assessing whether a patient is afflicted with prostate cancer;
CC        (b) monitoring the progression of prostate cancer in a patient;
CC        (c) assessing the efficacy of a test compound to inhibit prostate
CC        cancer in a patient;
CC        (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC        in a patient;
CC        (e) selecting a composition for inhibiting prostate cancer in a patient;
CC        (f) assessing the prostate cell carcinogenic potential of a compound;
CC        (g) determining whether prostate cancer has metastasized in a patient;
CC        (h) assessing the aggressiveness or indolence of prostate cancer in a
CC        patient;
CC        (I) is also useful as a pharmacodyanamic or pharmacogenomic marker.
XX
SQ        Sequence 691 BP; 186 A; 155 C; 199 G; 144 T; 7 other;

Alignment Scores:
Pred. No.:              3.36e-50        Length:              691
Score:                  666.50          Matches:             159
Percent Similarity:     73.82%          Conservative:        13
Best Local Similarity:  68.24%          Mismatches:          34
Query Match:            43.11%          Indels:              28
DB:                     23              Gaps:                8

US-10-054-313-1 (1-286) x ABV15988 (1-691)

Qy     24  GlyPheGlyMetPheTyrAlaValArgArgGlyArgGlyThrGlyValPheLeuThrThr   43
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3  GGGTTCGGGATGTTCTATGCCGTGAGGAGAGGGCCGCAAGACTGGGTCTTTCTGACCTGG   62

Qy     44  AsnGluCysArgArgAlaGlnValAspArgPheProAlaAlaArgPheLysPheAlaThr   63
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     63  AATGAGTGCAGAGCACAGGTGGACCGGTTTCCTGCTGCCAGATTTAAGAAGTTTGCCACA   122

Qy     64  GluAspGluAlaTrpGluAlaPheValArgLysSerAlaSerProGluValSerGluGlyHis   83
           ||||||||||||||||||||||||||||||||||||||||||||
```

WPI; 2000-293155/25.

DR
XX
PT Polynucleotide library comprising 1079 defined sequences, useful in
PT the form of an array to detect cancer or susceptibility to cancer -
XX
PS Claim 1; Page 233; 502pp; English.
XX
CC The present invention describes a library of polynucleotides comprising
CC 1079 nucleotide sequences (given in AAH30067 to AAH31145). Also described
CC are: (1) an isolated polynucleotide (I) having at least 90% identity to
CC one of the 1079 sequences; (2) a recombinant host cell containing (I);
CC (3) an isolated polypeptide (II) encoded by (I); (4) an antibody that
CC specifically binds to (II); (5) a vector comprising (I); and (6) a method
CC of detecting differentially expressed genes correlated with a cancerous
CC state of a mammalian cell comprising detecting a gene product encoded by
CC 65 of the 1079 sequences given in the specification. The polynucleotides
CC are used to monitor patients having (or susceptible) to cancer to detect
CC potentially malignant events at a molecular level before they are
CC detectable at a gross morphological level. The polynucleotides are also
CC useful for monitoring the efficacy of various therapies and preventive
CC interventions. Polynucleotide probes based on the disclosed sequences
CC are useful for chromosome mapping and detection of transcription levels.
CC The 1079 polynucleotide sequences were derived from a human colon cancer
CC cell line Km12L4-A cDNA library.
XX
SQ Sequence 389 BP; 78 A; 101 C; 130 G; 80 T; 0 other;

Alignment Scores:
Pred. No.:   Length: 389
Score: 1.8e-36   Matches: 508.50
Percent Similarity: 56.15% Conservative: 104
Best Local Similarity: 55.61% Mismatches: 1
Query Match: 32.89% Indels: 82
DB: 21   Gaps: 1

US-10-054-313-1 (1-286) x AAH30241 (1-389)

```
Qy      1 MetSerTrpLeuPheLeuAlaHisArgValAlaAlaLeuAlaAlaProCysArgArg  20
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     74 ATGAGCTGGCTTCTTTGTTCCTGGCCCACAGAGTCGCCTTGCCCTGGCCCTGCCGCCGC 133

Qy     21 GlySerArgGlyPheGlyMetPheTyrAlaValArgArgGlyArgLysThrGlyValPhe  40
            |||||||||||||||||||||||:::||||||||||||||||||||||
Db    134 GGCTCTCGCGGGTTCGGGATGTTCTATGCCGTGAGGAGGGGCCGCAAGACCGGGGTCTTT 193

Qy     41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArg-PheLysLy  60
            |||||:::||||||||||||||::::|||||||||||||||||||||
Db    194 CTGACCTGGAATGAGTGCAGAGCACAGGTGGACCGGTTTCCTGCTGCCAGACTTC----- 248

Qy     60 sPheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSe  80

Db    248 -------------------------------------------------------- 248

Qy     80 rGluGlyHisGluAsnGlnHisGlyHisGlnGluSerGluAlaLysProGlyLysArgLeuAr 100
            ::::|||||||||
Db    248 -------------------------------------------------------- 248

Qy    100 gGluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPr 120

Db    248 -------------------------------------------------------- 248

Qy    120 oSerValGluProAlaProProValSerArgArgAspThrPheSerTyrMetGlyAspPheVa 140
                                                                  ||
Db    249 -----------------------------------------------------GT 250

Qy    140 lValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgAlaGlyAlaGlyIl 160
            ||:::||||||||||||||||||||||:||||||||||||||||
Db    251 CGTCGTCACACTGATGGCTGCTGCTCCAGTAATGGGCGTAGAAGGGCCGCAGCAGAAT 310

Qy    160 eGlyValTyrTrpGlyProGlyHisProLeuAsnValGlyIleLeuProGlyArgGl 180
            |||||||||||||||||||||||||||||||||||||||||||||||
Db    311 CGGCGTTTACTGGGGGCCCGGGGCCATCTTTAAATGTAGGCATTAGACTTCCTGGGCGCA 370
```

RESULT 5
AAH30241
ID AAH30241 standard; cDNA; 389 BP.
XX
AC AAH30241;
XX
DT 27-JUL-2001 (first entry)
XX
DE Human colon cancer cell line Km12L4-A cDNA library derived sequence #175.
XX
KW Human; diagnosis; colon cancer; cancer; malignant; chromosome mapping;
KW detection; colon cancer cell line Km12L4-A; ss.
XX
OS Homo sapiens.
XX
PN WO200018916-A2.
XX
PD 06-APR-2000.
XX
PF 23-SEP-1999; 99WO-US22226.
XX
PR 28-SEP-1998; 98US-0102161.
PR 28-SEP-1998; 98US-0102180.
PR 29-SEP-1998; 98US-0102380.
PR 08-OCT-1998; 98US-0103815.
PR 27-OCT-1998; 98US-0105877.
XX
PA (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
PI Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
PI Lamson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
PI Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;
XX

```
Qy      180  nThrAsnGlnArgAlaGlu 186
              ||||||||||||||
Db      371  GACAAACCAAAGAGCGAA 389

RESULT 6
ABX41815
ID   ABX41815 standard; cDNA; 473 BP.
XX
AC   ABX41815;
XX
DT   20-FEB-2003  (first entry)
XX
DE   Bovine EST associated with lactation/muscle/fat deposition #6980.
XX
KW   Bovine; ss; EST; expressed sequence tag; lactation; LMFD;
KW   muscle deposition; fat deposition; genome mapping; gene identification;
KW   gene analysis; cattle breeding.
XX
OS   Bos Taurus.
XX
PN   US2002137139-A1.
XX
PD   26-SEP-2002.
XX
PF   24-SEP-2001; 2001US-0960352.
XX
PR   12-JAN-1999;  99US-115707P.
PR   11-JAN-2000; 2000US-0480902.
XX
PA   (BYAT/) BYATT J C.
PA   (MATH/) MATHIALAGAN N.
PA   (TAON/) TAO N.
PA   (WARR/) WARREN W C.
XX
PI   Byatt JC, Mathialagan N, Tao N, Warren WC;
XX
DR   WPI; 2003-110599/10.
XX
PT   New nucleic acid associated with lactation, and muscle and fat
PT   deposition, useful for genome mapping, gene identification and
PT   analysis, cattle breeding, or for genetically improving cattle  -
XX
PS   Claim 2; SEQ ID No 6980; 245pp; English.
XX
CC   The invention relates to a purified nucleic acid molecule associated with
CC   lactation or muscle and fat deposition (designated LMFD), derived
CC   from cattle, and the LMFD nucleic acid can specifically hybridise to a
CC   second nucleic acid molecule comprising any of 15112 nucleotide
CC   sequences, appearing as ABX34836-ABX49947, or complements of them.
CC   Also included are: (1) a transformed cell having a nucleic acid
CC   comprising an LMFD nucleic acid linked to a promoter and a 3' non-
CC   translated sequence that functions in the cell to cause termination of
CC   transcription and addition of polyadenylated ribonucleotides to a 3' end
CC   of the mRNA molecule; and (2) determining a level or pattern of a
CC   molecule in a bovine cell or tissue comprising: (a) incubating a marker
CC   nucleic acid (comprising any of the 15112 nucleic acid sequences or its
CC   complement or fragment) with a complementary nucleic acid molecule
CC   obtained from the bovine cell or tissue, where hybridisation between the
CC   marker nucleic acid and the complementary nucleic acid permits the
CC   detection of the molecule; and (b) detecting the level or pattern of the
CC   complementary nucleic acid, where the detection of the complementary
CC   nucleic acid is predictive of the level or pattern of the molecule.
CC   The LMFD nucleic acid is used for determining a level or pattern
CC   of a molecule in a bovine cell or tissue. It is useful for genome
CC   mapping, gene identification and analysis, cattle breeding, preparation
CC   of constructs for use in cattle gene expression, or for genetically
CC   improving cattle. The present sequence is one of the 15112 bovine
CC   LMFD EST (expressed sequence tag) nucleic acids.
CC   Note: The present sequence was not shown in the specification but
CC   was obtained in electronic format from the USPTO web site:
CC   seqdata.uspto.gov/sequence.html?DocID=20020137139.
XX
SQ   Sequence 473 BP; 129 A; 109 C; 140 G; 94 T; 1 other;
```

```
Alignment Scores:
Pred. No.:                   9.51e-30            Length:         473
Score:                      434.00              Matches:         79
Percent Similarity:          88.78%             Conservative:     8
Best Local Similarity:       80.61%             Mismatches:      11
Query Match:                 28.07%             Indels:          25
                                                Gaps:             0

US-10-054-313-1 (1-286)  x ABX41815 (1-473)

Qy      189  AlaAlaAlaCysLysAlaIleGluGlnAlaLysThrGlnAsnIleAsnLysValLeuTyr  208
                 ||||||||||||||||||||||||||||||||||||||:::||||
Db       79  GCAGCTGCAAGCCATAGAGCAACTAAGCTCAGGACATCACTAAGCTGGTTCCTAC  138

Qy      209  ThrAspSerMetPheThrIleAsnGlyIleThrAsnTrpValGlnGlyTrpLysLysAsn  228
                 ||||||||||||||||||||||||||||||||||||:::||||:::|||
Db      139  ACAGACAGCATGTTTACCATCAATGGCATCACCACCTGGGTGAAGGGCTGGAAGCAGAAT  198

Qy      229  GlyTrpLysThrSerAlaGlyValLysGluValIleAlaGlyIleAspPheValAlaLeuGlu  248
                 ||:::|||||:::|||||||||||||||||||||||||||||
Db      199  GGCTGGAGGACCAGCACCGGAAGAGGTGACCACCAAGAGGACTTCGGGAGCTGGAG  258

Qy      249  ArgLeuThrGlnGlyMetAspIleGlnTrpMetHisValProGlyHisSerGlyPheIle  268
                    |:::|||||||||||||||||||||||||||||||||||||||||
Db      259  CGGCTGGCGCGCGGGCCATGGACATTCAGTGGATGCACGTTCCTGGCCATTCGGGGTTTAAA  318

Qy      269  GlyAsnGluGlyAlaAspArgLeuAlaArgGluValAlaLysGlnSerGlyGluAsp  286
                 ||||||||||||||||||||||||||||||||||||||||||||||
Db      319  GGCAATGAAAGGCGGACAACGTCGAGAGAGAGGCTCGAAACACTCCGCAGAC  372
```

```
RESULT 7
ABL03247
ID   ABL03247 standard; cDNA; 1071 BP.
XX
AC   ABL03247;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 4223.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB59144.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 4223; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
```

```
Db    823  AAGAACAATCAGCCTGGTTAAAACGTTGTTGACTCGATTCCAAGGAACTGGATAAACTGCTTCAG  882

Qy    253  ------GlyMetAspIleGlnTrpMetHisValProGlyHisSerGlyPheIleGlyPheAsn  270
                 :::::|||  |||::||:|
Db    883  GAGAACAACATCACCGTGAAATGGAACTATGTGGAGGCCCACAAGGGCATAGAGGGCAAT  942

Qy    271  GluGluLeuAspAlaArgLeuAlaArgGluGlyAla  281
                |||  |||||:||| ||| |:|| |:::|||
Db    943  GAAATGGCAGACAAATTGGCGCGGCAAGGATCC  975
```

```
RESULT 8
ABX38735
ID   ABX38735 standard; cDNA; 407 BP.
XX
AC   ABX38735;
XX
DT   20-FEB-2003  (first entry)
XX
DE   Bovine EST associated with lactation/muscle/fat deposition #3900.
XX
KW   Bovine; ss; EST; expressed sequence tag; lactation; LMFD;
KW   muscle deposition; fat deposition; genome mapping; gene identification;
KW   gene analysis; cattle breeding.
XX
OS   Bos Taurus.
XX
PN   US2002137139-A1.
XX
PD   26-SEP-2002.
XX
PF   24-SEP-2001; 2001US-0960352.
XX
PR   12-JAN-1999;  99US-1157Q7P.
PR   11-JAN-2000; 2000US-0480902.
XX
PA   (BYAT/) BYATT J C.
PA   (MATH/) MATHIALAGAN N.
PA   (TAON/) TAO N.
PA   (WARR/) WARREN W C.
XX
PI   Byatt JC, Mathialagan N,   Tao N,   Warren WC;
XX
DR   WPI; 2003-110599/10.
XX
PT   New nucleic acid associated with lactation, and muscle and fat
PT   deposition, useful for genome mapping, gene identification and
PT   analysis, cattle breeding, or for genetically improving cattle
XX
PS   Claim 2; SEQ ID No 3900; 245pp; English.
XX
CC   The invention relates to a purified nucleic acid molecule associated with
CC   lactation or muscle and fat deposition (designated LMFD), derived
CC   from cattle, and the LMFD nucleic acid can specifically hybridise to a
CC   second nucleic acid molecule comprising any of 15112 nucleotide
CC   sequences, appearing as ABX34836-ABX49947, or complements of them.
CC   Also included are; (1) a transformed cell having a nucleic acid
CC   comprising an LMFD nucleic acid linked to a promoter and a 3'-non-
CC   translated sequence that functions in the cell to cause termination of
CC   transcription and addition of polyadenylated ribonucleotides to a 3' end
CC   of the mRNA molecule; and (2) determining a level or pattern of a
CC   nucleic acid in a bovine cell or tissue comprising: (a) incubating a marker
CC   nucleic acid (comprising any of the 15112 nucleic acid sequences or its
CC   complement or fragment) with a complementary nucleic acid molecule
CC   obtained from the bovine cell or tissue, where hybridisation between the
CC   marker nucleic acid and the complementary nucleic acid permits the
CC   detection of the molecule; and (b) detecting the level or pattern of the
CC   complementary nucleic acid, where the detection of the complementary
CC   nucleic acid is predictive of the level or pattern of the molecule.
CC   The LMFD nucleic acid is used for determining a level or pattern
CC   of a molecule in a bovine cell or tissue. It is useful for genome
CC   mapping, gene identification and analysis, cattle breeding, preparation
CC   of constructs for use in cattle gene expression, or for genetically
CC   improving cattle. The present sequence is one of the 15112 bovine
```

```
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1071 BP; 320 A; 229 C; 292 G; 230 T; 0 other;

Alignment Scores:
Pred. No.:                                 Length:              1071
Score:               3.03e-28              Matches:             107
Percent Similarity:  422.00                Conservative:        31
Best Local Similarity: 44.37%              Mismatches:          113
Query Match:         34.41%                Indels:              60
DB:                  27.30%                Gaps:                7
                     23

US-10-054-313-1 (1-286) x ABL03247 (1-1071)

Qy     28  PheTyrAlaValArgArgGlyLysThrGlyValPheLeuThrTrpAsnGluCysArg  47
               |||  |||     || :|| ||| ||||
Db     52  TTTTACGCTGTAGCCAGTGGGCGGCCGGGTCTGGAGCTCTACGGCGTTCTTGGCAGAGTGCCAG  111

Qy     48  AlaGlnValAspArgPheProAlaAlaArgPheLysLysPheAlaThrGluAspGluAla  67
               |||||:|||     || ||||| ::|:|| :|||||||
Db    112  GAGCAGGTCAAGGGTTCAAGACACGCCAAATACAACAAGAGTTTAAGACACGCCAGGAAGCG  171

Qy     68  TrpAlaPheValArg------LysSerAlaSerProGlu--  78
               ||||| ||| |:::|||
Db    172  GATCAGTTCGTTAATGGCTGCAAGTCGTATGCTCCGGAGGATGTGGCCGCTGCGGC  231

Qy     78  ------------------------------------------  78
Db    232  AAGGAAAAAGGCGTCACTGGCCAGTTGGAAAAACAGTATCGAAGTGATAAGAACCCCAAG  291

Qy     79  ---ValSerGluGlyHisGlnHisGly------------------------  88
                |||  |||||
Db    292  TACACAGATGAGTGCCCGAAGAGATCACGACCTAGCCGAGGATGATCTGAATGCTGCC  351

Qy     89  -----------GlnGluSerGluLysLysPro----------  95
                 :|| |||
Db    352  ATGAACGAGGTTGAGGAGGATCCAAGGCCATCTAACAGTAGCAATTTACCAGACATTCTC  411

Qy     96  GlyLysArgLeuArgGluProLeuAspGlyHisGluSerGlyHisGluSerAlaGlnProTyrAla  115
               :|| |||         ||| ::|:: ||
Db    412  AATCGCAGCGGAGGGCACAACCAGCCGCGATAAGCGGAACAAGATCCCACGTCATGCC  471

Qy    116  LysHisMetLysProSerValGluProAlaProProValSerArgArgAspThrPheSerTyr  135
                                                   |||  ::|
Db    472  TCTCAGGTCTCA----------GAGCCACCAGGACTCAGCAAGTGGGTGCCTTCCAGTTT  522

Qy    136  MetGlyAsp----------PheValValTyrThrAspGlyCysCysSerSerAsnGly  152
                :::::|:: ||:| :|||  |||  |||||
Db    523  GAAATCGATGCAGAGGGCTATGTCATTGTGTACACCAGATGGCTTCCTGTACACCGGA  582

Qy    153  ArgArgLysProArgAlaGlyIleTyrValGlyValTyrTrpGlyHisProLeuAsnVal  172
               :|| |||||| ||| |: |:  ||| |:||| |||
Db    583  CGAGCCGGAGCCGTGCCGGCTATGCGGTTATTTCGGCAAGAATCACCAGCTAAACGACA  642

Qy    173  GlyIleArgLeuProGlyArgGlnThrAsnGlnArgAlaGlyIleHisAlaAlaCysLys  192
               |||||:|||  |:|||:||  |||||
Db    643  GCCAAGCCCGTGGAAGAGCGACGCGGTTACAAATAATGTAGGTGAAATACAAGCGGCCATTCAT  702

Qy    193  AlaIleGluGlnIleAlaLysThrGlnAsnIleAsnLysLeuValLeuTyrThrAspSerMet  212
               |||||::|:  |||  ||| |||  |::|| :|:|||
Db    703  GCCATTAAAACAGCTCTTGACTTGGAATACAGAAGCTGTGCATCAGCAGACTCTCGAG  762

Qy    213  PheThrIleAsnGlyIleThrAsnTrpValGlnGlyTrpLysLysAsnGlyTrpLysThr  232
               :::::| || |:::|| |||  ||||:|| ||||||
Db    763  TTTTGATCAACTCCATAACGCTGTGCGTGGGTTGGAAAAAAGGGATTGGAAGCTA  822

Qy    233  SerAlaGlyLysGluValIleAsnLysGluAspPheValAlaLeuArgGluLeuThrGln  252
               :: |||     |||  |||  |::::|||
```

CC　LMFD EST (expressed sequence tag) nucleic acids.
CC　Note: The present sequence was not shown in the specification but
CC　was obtained in electronic format from the USPTO web site:
CC　seqdata.uspto.gov/sequence.html?DocID=2002037139.
XX
SQ　Sequence 407 BP; 97 A; 95 C; 144 G; 71 T; 0 other;

Alignment Scores:
Pred. No.:　　　　　　　1.02e-26
Score:　　　　　　　　　399.00　　　　　Length:　　　　　407
Percent Similarity:　　75.81%　　　　　Matches:　　　　　81
Best Local Similarity: 65.32%　　　　　Conservative:　　13
Query Match:　　　　　　25.81%　　　　　Mismatches:　　　24
DB:　　　　　　　　　　　25　　　　　　　Indels:　　　　　6
　　　　　　　　　　　　　　　　　　　　　Gaps:　　　　　　2

US-10-054-313-1 (1-286) x ABX38735 (1-407)

Qy　　1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaAlaAlaLeuProCysArgArg 20
Db　 52 ATGAGCCCGCTGCTGGGCGTGGTGCACAGAGTGGCTTTGGCCTTTGTGGGATGC------ 105

Qy　 21 GlySerArgGlyPheGlyMetPheTyrAlaValArgArgGlyArgLysThrGlyValPhe 40
Db　106 ------CGCAGTCTGCCATGTTCTATGCCGTGAGGAGGGGCCCAAGGCCGGGGTCTTC 159

Qy　 41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys 60
Db　160 CTGACCTGGGATGAATGCAGAGCACCAGGTGGACCGCTTTCCTGCAGCCAGATTTAAGAAG 219

Qy　 61 PheAlaThrGluAspGlyAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80
Db　220 TTCGCCACAGAGGACGGAGCCTGGGCCTTTGTCAGGAAGTCTGAAAGCCCCAGTGATTCA 279

Qy　 81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100
Db　280 GAAGGGCAAAAACTAAACACGTAGAAGAACCCAAGGCGAAAGCGAGCAAGCGACTCCGT 339

Qy　101 GluProLeuAspGlyGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120
Db　340 GAGCCCTTGGATGAAGAT------GAAGATGCAAAGGCCTGTGCCGAAGCACGTGCGGGAG 393

Qy　121 SerValGluPro 124
Db　394 AGCGGCAGAGCCA 405

RESULT 9
ABS77198
ID　ABS77198 standard; cDNA; 764 BP.
XX
AC　ABS77198;
XX
DT　12-DEC-2002 (first entry)
XX
DE　Frog embryonic gene sequence Q9925613.
XX
KW　Frog; ss; embryonic development; developmental disorder;
KW　microarray; cell differentiation.
XX
OS　Xenopus laevis.
XX
PN　US2002081610-A1.
XX
PD　27-JUN-2002.
XX
PF　23-JUL-2001; 2001US-0910943.
PR　21-JUL-2000; 2000US-219658P.
XX
PA　(UYRQ) UNIV ROCKEFELLER.
XX
PI　Hemmati-Brivanlou A, Altmann CR;
XX
DR　WPI; 2002-626534/67.

XX
PT　Nucleic acid array containing Xenopus embryonic nucleic acids is useful
PT　to identify genes involved in embryonic development, to identify
PT　different types of embryonic cells, and to diagnose developmental
PT　disorders.
XX
PS　Claim 1; Page 468; 823pp; English.
XX
CC　The invention relates to a nucleic acid array, where each coordinate
CC　contains a single nucleic acid species having one of 770 nucleotide
CC　sequences (appearing as ABS7747-ABS77516) a of a Xenopus embryonic gene
CC　product, or its complement or hybridisable fragment of not less than
CC　20 contiguous nucleotides of one of those sequences. Also included are
CC　detecting differential expression of embryonic genes, comprising:
CC　(a) contacting a nucleic acid array comprising genes expressed in
CC　embryonic but not mature cells with nucleic acids from sample and
CC　control cells; and (b) detecting differential hybridisation of nucleic
CC　acids from the sample cells relative to the control cells; and detecting
CC　defects in development, comprising: (a) contacting nucleic acids from
CC　test cells undergoing development with a nucleic acid array of gene
CC　products known to play a fundamental role in the development process; and
CC　(b) detecting a difference in expression of a fundamental gene in the
CC　sample cells relative to a standard. The invention is useful to identify
CC　genes involved in embryonic development and related processes such as
CC　cell differentiation. This would be useful for diagnosing developmental
CC　disorders and for identifying different types of embryonic cells.
CC　The present sequence is one of the 770 Xenopus embryonic cDNA sequences.
XX
SQ　Sequence 764 BP; 212 A; 160 C; 189 G; 192 T; 11 other;

Alignment Scores:
Pred. No.:　　　　　　　2.84e-23
Score:　　　　　　　　　364.00　　　　　Length:　　　　　764
Percent Similarity:　　58.38%　　　　　Matches:　　　　　83
Best Local Similarity: 47.98%　　　　　Conservative:　　18
Query Match:　　　　　　23.54%　　　　　Mismatches:　　　57
DB:　　　　　　　　　　　24　　　　　　　Indels:　　　　　15
　　　　　　　　　　　　　　　　　　　　　Gaps:　　　　　　5

US-10-054-313-1 (1-286) x ABS77198 (1-764)

Qy　 27 MetPheTyrAlaValArgArgGlyArgLysThrGlyValPheLeuThrTrpAsnGluCys 46
Db　249 ATGTTTTACGCTGTAAGGACTGCCGTAAGCCTGGAGTCTCAACAATACGTGGGATGAATGT 308

Qy　 47 ArgAlaGlnValAspArgPheProAlaAlaArgPheLysLysPheAlaThrGluAspGlu 66
Db　309 AAAGAGCAAGTGGATCGATTTCCTTTAGCAAGTACAGAAGTTTGCTCAGAGGAAGAGAT 368

Qy　 67 AlaTrpAlaPheValArg------LysSerAlaSerProGluValSerGluGlyHisGlu 84
Db　369 GCCTGGGAATTTGTGAGGAACACTCAGGAATCATCATCGAAAGGTTCTACTAGTGTGAA 428

Qy　 85 AsnGlnHisGlyGlnGluSerGluAlaLys------------ProGly 96
Db　429 ACAAAGGAGCCTCTACACAAGCTACAAAAGCTGCAGGACTGCCATAATGTCATACCTCAG 488

Qy　 97 LysArgLeuArgGluPro---LeuAspGlyAspGlyHisGluSerAlaGlnProTyrAla 115
Db　489 TCCAGAAGAAGAGACCACTACTACAGAGCTCAAGCACTGAGAAAGCATCCTCACCTAAA 548

Qy　116 LysHisMetLysProSerValGluProAlaProProValSerArgAspThrPheSerTyr 135
Db　549 AGAAGCAAGCTCATTGATATCACTGATTTACCATTCGTCACATAATGGAACTTTTACCTAC 608

Qy　136 MetGlyAspPheValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLys 155
Db　609 ATGGGAGACGCTGCTGTGTATACACTGATGGCTGCTGCTAGCGGGAATGGCCGGTAAA 668

Qy　156 ProArgAlaGlyLeuGlyValTyrTrp-GlyProGlyHisProLeuAsnValGlyIleAr 175
Db　669 GCACGAGCTGGTATAGGTGTATACTGGGGCCAAAGGCCTTCCTCTNAACCTTGCAGAAA 728

Qy　175 gLeuProGlyArgGlnThrAsn------GlnArgAla 185

```
Db      729 GCTT---GGAAGGGAGGCAAACTTAACCCCAGGCGGGCT 762
            :.|||  |||||:.:||      ||||||||

RESULT 10
ABL03246
ID   ABL03246 standard; cDNA; 3290 BP.
XX
AC   ABL03246;
XX
DT   26-MAR-2002 (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 4220.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB59143.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 4220; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 3290 BP; 830 A; 751 C; 867 G; 842 T; 0 other;
```

Alignment Scores:
```
Pred. No.:                              Length:       3290
Score:             7.998-22             Matches:       122
Percent Similarity: 37.41%              Conservative:   31
Best Local Similarity: 29.83%           Mismatches:    125
Query Match:       23.06%               Indels:        132
Db:                23                   Gaps:           12
```

```
US-10-054-313-1 (1-286) x ABL03246 (1-3290)

Qy     4 LeuLeuPheLeuAlaHisArgValAlaLeuAlaAlaLeuPro--------CysArgArg 20
           :::   |||| :.:||      ||||||  |  ||||| :: 
Db   971 ATATGTTTCGATAGACAACGTAAGAAAATATGTTACTTCCGCCGTATTTTTGCTGGAAC 1030

Qy    21 GlySerArgGlyPheGlyMetPheTyrAlaValArgArgGlyArgLysThrGlyValPhe 40
          |||  ||:|||||||  |||||:.:||||||:||| |||:|||||| ||||||:||||| 
Db  1031 CTCCAACGCTGCACGATGGCGTTTACGCTGTAGCCAGTGGCGGCGGTCTCGGAGTCTAC 1090

Qy    41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaArgPheLysLys 60
```

```
Db  1091 GGCTCTTGGCGAGTGCAGGAGGAGGCGGTCAAGGGTTTCAAGAACGCCAATACAAGAAG 1150
             :.|||         |||||||              ||||||   :::  :|||||||

Qy    61 PheAlaThrGluAspGluAlaTrpAlaPheValArg-----LysSerAlaSerProGlu 78
          |||           |||  ||||  |||   ||||||       |||||: ::|||    
Db  1151 TTTAAGACACCCCAGGAAGCGGATCAGTTCGTTAATGGCTGCAAGTCGTATGCTCCGCAG 1210

Qy    78 ---------------------------------------- 78

Db  1211 GATGTGGCCGTGCCGCTGGGCAAGGAAAAGGCGTCACTGGC CAGTTGGAAAAACAGTATC 1270

Qy    79 --------ValSerGluGlyHisGluAsnGlnHis------ 87
                   |||      ||    |||
Db  1271 GAAGTGAATAGAACCCAAGTACACAGATGAGTGGCCCGAGAGGATCACGACCTAGCC 1330

Qy    88 ------------GlyGlnGlu-------------------- 90
                      |||

Db  1331 GAGGATGATCTGGTAGGCATGAATATACATACGTTTGCAAAGTCCTGACCTTTCATCCTT 1390

Qy    91 SerGluAlaLysProGlyLysArgLeuArgGlu------ 101
          |||||||           ||||| 
Db  1391 TCAGAATGCTG-CCATGAACCAGGTTGAGGGAGATCCAAAGCCATCTAACAGTAGCAATT 1449

Qy   102 ----------ProLeuAspGlyAspGlyHisGluSerAlaGlnProTyr 114
                     :.:::|||          
Db  1450 TAGTATGAGACCCTAGTACATACTGTAGAAAAAGTTAAACATAGCTTCTTTCAGCCAGAC 1509

Qy   115 ----------ValGlu 123
                     |||
Db  1510 ATTCTCAATCGCAAGCGGAAGGGCCACAACCAGCGGCGATAAGCGGAACAAGATCCCACGT 1569

Qy   124 ----------PheSerTyrMet 136
                     :::
Db  1570 CATGCCCTCTCAGGTCTCAGAAGCACAGGAGTCAAGCAAGGTGGGTGCCTTCCAGTTTGAA 1629

Qy   137 GlyAsp----------PheValValTyrThrArgGlyCysCysSerSerAsnGlyArg 153
          ||| |||         :.:::|||
Db  1630 ATCGATGCGAGAGGCTATGTCATTGTGTACACAGATGGCTCCTGCATAGGCAACGGACGA 1689

Qy   154 ArgLysProArgAlaGlyIleGlyValValTyrProGlyValTyrTrpGlyGlyHisProLeu--- 170
          |||
Db  1690 GCCGGAGCCTGTGCCGGCTATGGCGTTTATTCGGCAAGAATCACCAGCTGTAAGTGGTA 1749

Qy   171 -----------------------------------------AsnValGlyIleAr 175
                                                     |||
Db  1750 ACTGCAGCACAATAACTTGCTTTAACCTATATGTTCTTTGCAGAAACGAGCCAAGCC 1809

Qy   175 gLeuProGlyArgGlnThrArgAsnGlnArgAlaGluHisAlaAlaCysLysAlaIleGl 195
          |||
Db  1810 CGTGGAAGGACGCGGTTACAAATAATGTAGGTGAAATACAAGCGGCCATTCATGCCATTAA 1869

Qy   195 uGlnAlaLysThrGlnAsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIl 215
          :
Db  1870 AACAGCTCTTGACTTGGAATACACAGAGCTGTGCATCAGCACAGACTCTCAGTTTTGAT 1929

Qy   215 eAsnGlyIleThrAsnTrpValGlnGlyTrpValGlnGlyTrpLysThrSerAlaGl 235
          |||
Db  1930 CAACTCCATAACGCTGTGGGTTGCGGGTGGAAAAAAGGGATTGGAAGCTAAAGACAA 1989

Qy   235 yLysGluValIleAsnLysGluAspPheValAlaLeuGluArgLeuThrGln-----Gl 253
          :::|||
Db  1990 TCAGCCTGTTAAAAACGTTGTTGACTTCAAGGAACTGGATAAACTGCTTCAGGAGAACAA 2049

Qy   253 yMetAspIleGlnTrpMet------------------- 259
          ::: :.:::|||
Db  2050 CATCACCGTGAAATGGGTGCGTAACTAAATTTAAATAAGGTCCTGGAAAATACTTGACTT 2109

Qy   260 -----------HisValProGlyHisSerGlyPheIleGlyAsnGluGluA1 273
                     :::|||
Db  2110 CAGTTAATATCCCCAGAACTATGTGGAGCCCACCAAGGCCATAGAGGGCAATGAAATGGC 2169

Qy   273 aAspArgLeuAlaArgGluGluGlyAla 281
          |||:::|||||||||||:.:::
```

```
Qy       184  ArgAlaGluIleHisAlaAlaCysLysAlaIleGluGlnAlaLysThrGlnAsnIleAsn  203
              |||  |||  |    ||||||  |  ::|  :::||:::::||  |     |||
Db     45032  CGCATGGAATTGATGGCTGCCATCACAGCTTTGGAGGCAACGCCTGCACAG-----ATT  44979

Qy       204  LysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsnTrpValGln  223
              ::|      ||||:  |    |   ||  ||  ::||:||  :::|:
Db     44978  CCTCTGCAACTTGGACAGATTCAGGCTATCGTATCTGTTTAAAGATGGCATAACTCAGTGGATTGGC  44919

Qy       224  GlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyValIleAsnGlyGluAsp  243
              ||||||  |||  |||||  ||  :::  |||  :  |||
Db     44918  GGTTGGAAGTTGCGTGGTTGGAAAAAAGCAGATGGCAAGCCTGTCTTAAATCAAGACCTA  44859

Qy       244  PheValAlaLeuGluArgLeuThrGlnMetAspIleGlnTrpMetHisValProGly  263
              |||:|:::|||:||||||
Db     44858  TGGCAACGATTGGATCAGCTGACCCAAATCGCATCATTGATTGGCAATGGATCAAAGGC  44799

Qy       264  HisSerGlyPheIleGlyLysAsnGluAlaAspArgLeuAlaArgGluGly  280
              :::  |||||  |  |  :||||||||
Db     44798  CATGCTGGTCATCGTCTGCAATGAGATGGCAGACCACTGGCCAATAAAGGC  44748
```

```
RESULT 12
AAT24248
ID   AAT24248  standard; cDNA to mRNA; 378 BP.
XX
AC   AAT24248;
XX
DT   16-OCT-1996  (first entry)
XX
DE   Human gene signature HUMGS06264.
XX
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
XX
OS   Homo sapiens.
XX
PN   WO9514772-A1.
XX
PD   01-JUN-1995.
XX
PF   11-NOV-1994;    94WO-JP01916.
XX
PR   12-NOV-1993;    93JP-0355504.
XX
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
XX
PI   Matsubara K,  Okubo K;
XX
DR   WPI; 1995-206931/27.
XX
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
XX
PS   Claim 1; Page 1565; 2245pp; Japanese.
XX
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in AAT19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
```

```
Db     2170  AGACAAATTGGCGCGGCAAGGATCC  2194
```

```
RESULT 11
AAF28553/c
ID   AAF28553  standard; DNA; 119211 BP.
XX
AC   AAF28553;
XX
DT   04-APR-2001  (first entry)
XX
DE   Genomic fragment #40.
XX
KW   Genomic library; bacteria; human upper airway; otitis media; sinusitis;
KW   bronchopulmonary; endocarditis; meningitis; ss.
XX
OS   Moraxella catarrhalis.
XX
PN   WO200078968-A2.
XX
PD   28-DEC-2000.
XX
PF   16-JUN-2000;  2000WO-US16649.
XX
PR   18-JUN-1999;   99US-0140121.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Lagace RE,  Patterson C,  Berg KL;
XX
DR   WPI; 2001-041427/05.
XX
PT   Genomic library for identifying diagnostic and therapeutic
PT   compositions, and for identifying virulence factors, regulatory
PT   elements and drug targets, comprises Moraxella catarrhalis nucleic
PT   acids -
XX
PS   Claim 1; Page 459-485; 545pp; English.
XX
CC   The present invention relates to a Moraxella catarrhalis genomic library
CC   comprising of a combination of 41 nucleic acid molecules (see
CC   AAF28514-AAF28554). The library has a number of uses described in the
CC   specification e.g. is useful for aidentifying diagnostic and therapeutic
CC   compositions. M.catarrhalis (Branhamella catarrhalis) is a large
CC   aerobic, gram-negative diplococcus, normally found among the bacterial
CC   flora of human upper airways. M.catarrhalis is known to cause acute,
CC   localised infections such as otitis media, sinusitis and bronchopulmonary
CC   infection and life-threatening, systemic diseases including endocarditis
CC   and meningitis.
XX
SQ   Sequence 119211 BP; 34729 A; 23773 C; 25664 G; 35041 T; 4 other;
```

```
Alignment Scores:
Pred. No.:              2.15e-12
Score:                  272.00
Percent Similarity:     52.23%         Length:        119211
Best Local Similarity:  39.49%         Matches:          62
Query Match:            17.59%         Conservative:     20
                                       Mismatches:       51
DB:                                    Indels:           24
                                       Gaps:              5
```

```
US-10-054-313-1 (1-286) x AAF28553 (1-119211)

Qy       136  MetGlyAspPheValValValTyrThrAspTyrCysCysSerSerAsnGlyArgArgLys  155
              |||                          ||||||||||||  ::::::
Db     45182  ATGAGCCAACTTAATATCGCTTATACCGACGGCGCCCTGTAAAGGCAATGGAAAAACAAGGC  45123

Qy       156  ProArgAla---GlyIleGlyValTyr----------------------  163
              |||                ::  |||||
Db     45122  GTATCTGCAGGCGGTTGGGCGGTGTATTCGATTATTTCATTATTTCAATGGCGATGAGCGGCATCTG  45063

Qy       164  TrpGlyProGlyHisProLeuAsnValGlyIleLeuProGlyAspGlyArgGlnThrArgGlnGln  183
              |||||||                                      ||||||
Db     45062  TGGGGC--GGTGAGCCT----------------------GATACGACCAATAAT  45033
```

```
XX
SQ   Sequence 378 BP; 98 A; 61 C; 106 G; 94 T; 19 other;

Alignment Scores:
Pred. No.:                                Length:          378
Score:                  5.46e-13          Matches:          47
Percent Similarity:     244.00            Conservative:      0
Best Local Similarity:  97.92%            Mismatches:        1
Query Match:            97.92%            Indels:            0
DB:                     15.78%            Gaps:              0
                        16

US-10-054-313-1 (1-286) x AAT24248 (1-378)

Qy  239 IleAsnLysGluAspPheValAlaLeuGluArgThrGlnGlyMetAspIleGlnTrp 258
            |||||||||||||||||||||||||||||||||||||||||||||||||
Db    2 ATCAACAAAGAAGACTTTGGCACTGGAGAGGCTTACCCAGGGGATGGACATTCAGTGG 61

Qy  259 MetHisValProGlyHisSerGlyPheIleGlyGlyAsnGluGluAlaAspArgLeuAlaArg 278
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   62 ATGCATGTTNCTGGTCATTCGGGATTTATAGGCAATGAAGAGCTGACAGATTAGCCAGA 121

Qy  279 GlyGlyAlaLysGlnSerGluAsp 286
            ||||||||||||||||||
Db  122 GAAGGAGCTAAACAATCGGAAGAC 145

RESULT 13
AAX20568/c
ID   AAX20568  standard; DNA; 7874 BP.
XX
AC   AAX20568;
XX
DT   05-MAY-1999  (first entry)
XX
DE   Polynucleotide sequence from the genome of Treponema pallidum.
XX
KW   Treponema pallidum infection; syphilis; Borrelia infection; animal;
KW   enzyme production; ds.
XX
OS   Treponema pallidum.
XX
XX
PN   WO9959034-A2.
XX
PD   30-DEC-1998.
XX
PF   23-JUN-1998;  98WO-US13041.
XX
PR   24-JUN-1997;  97US-0050667.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Fraser CM;
XX
DR   WPI; 1999-081273/07.
XX
PT   New isolated Treponema pallidum nucleic acids - used to develop
PT   products for the detection, diagnosis, characterisation, prevention
PT   and therapy of T. pallidum infections, particularly syphilis
XX
PS   Claim 1; Page 560-565; 1150pp; English.
XX
CC   AAX20500-21243 represent polynucleotide sequences from the genome of
CC   Treponema pallidum. The sequences can be used for detection,
CC   diagnosis, characterisation, prevention and therapy for T. pallidum
CC   infections, particularly syphilis. They can also be used for detecting
CC   diseases related to Borrelia infections in animals, and for the
CC   production of biosynthetic products such as enzymes.
XX
SQ   Sequence 7874 BP; 1590 A; 1949 C; 2387 G; 1929 T; 19 other;

Alignment Scores:
Pred. No.:                                Length:         7874
Score:                  6.43e-11          Matches:          52
Percent Similarity:     239.00            Conservative:     25
                        46.67%
```

```
Best Local Similarity:  31.52%          Mismatches:       58
Query Match:            15.46%          Indels:           30
DB:                                     Gaps:              2
                        20

US-10-054-313-1 (1-286) x AAX20568 (1-7874)

Qy  140 ValValValTyrThrAspGlyCysCysSerSerAsn--------------------- 151
            ::: :||::|||||::||||||||
Db 5784 CTTACCCTGTACACTGACGGGCCCTGTCTGGGAAACCCGGTCCGGCGGTGGGCGTT 5725

Qy  152 ----------------GlyArgArgLysProArgAlaGly 159
                            ||||  |||||||
Db 5724 GCTCTGTGCCCTCAGACGTACCCTTTCTAGAAACCCACCAGAAGCTGCC 5665

Qy  160 IleGlyValTyrTrpProGlyHisProLeuAsnValGlyIleArgLeuProGlyArg 179
            |||||:||::||||||||||
Db 5664 GCATTCACTAGAAGCGGCAGCGCCTACCCG--------------------------- 5635

Qy  180 GlnThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaIleGluGlnAlaLysThr 199
            ||||| ||::|||||||||||
Db 5634 AGCACAAACAATCGCATGGAACTGTGCGGTCATCAACGACTGCAAGAGCACACGGA 5575

Qy  200 GlnAsnIleAsnLysGluValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThr 219
            :::  ||:::|||:::
Db 5574 CCGCGTCGCAGAGGCGGTGGTAGTGGTCACCGGCAAAAGGCATCACC 5515

Qy  220 AsnTrpValGlnGlyTrpLysLysAsnGlyTyrTrpLysSerArgAlaGlyLysLysGluValIle 239
            |||:::
Db 5514 CAGTGGATCCATTACCTGAAACACAACGGCTGCCAAACAACCGGTCAAG 5455

Qy  240 AsnLysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMet 259
            ||||:::
Db 5454 AACAAGGATTTGTGGGAAGCCACTCGCCGACGCCCTATCCGTCGAGTGGCGC 5395

Qy  260 HisValProGlyHisSerGlyPheIleGlyGlyAsnGluGluAlaAspArgLeuAlaArgGlu 279
            |||::::||:::
Db 5394 TGGGTAAAAGGACATGCAGGAGACCCCTACACGACTGTGCGACCGGCTAGCCACCGAC 5335

Qy  280 GlyAlaLysGlnSer 284
            |||:::::::
Db 5334 GCTGCGCGCCGCGCT 5320

RESULT 14
AAF94380
ID   AAF94380  standard; DNA; 462 BP.
XX
AC   AAF94380;
XX
DT   04-JUN-2001  (first entry)
XX
DE   Haemophilus influenzae essential bacterial gene SEQ ID NO:71.
XX
KW   Haemophilus influenzae; essential bacterial gene; identification;
KW   otitis media; meningitis; upper respiratory tract infection;
KW   infection; antimicrobial; ds.
XX
OS   Haemophilus influenzae.
XX
XX
PN   WO200111033-A2.
XX
PD   15-FEB-2001.
XX
PF   03-AUG-2000; 2000WO-US21176.
XX
PR   04-AUG-1999;  99US-0368382.
XX
PA   (ABBO ) ABBOTT LAB.
XX
PI   Chovan LE, Hessler PE, Reich KA;
XX
DR   WPI; 2001-147511/15.
DR   P-PSDB; AAB8527.
XX
```

PT Essential bacterial genes from Haemophilus influenzae and methods for
PT identifying 'essential' genes that may be potential therapeutic targets
PT -
XX
PS Claim 2; Page 116-117; 185pp; English.
XX
CC AAF94345 to AAF94409 represent essential bacterial genes from
CC Haemophilus influenzae, which encode the proteins given in AAB88492 to
CC AAB88556. The present invention also describes methods for identifying
CC essential bacterial genes (i.e. those essential to the survival of a
CC bacterium) using a transposition system. The methods are used to
CC identify essential genes from bacteria, especially H. influenzae (which
CC causes otitis media, meningitis and upper respiratory tract infections)
CC which may be used as targets for potential antimicrobial agents.
CC AAF94410 to AAF94416 represent PCR primers used in the exemplification
CC of the present invention.
XX
SQ Sequence 462 BP; 180 A; 74 C; 96 G; 111 T; 1 other;

Alignment Scores:
Pred. No.:   3.25e-12   Length: 462
Score: 236.50   Matches: 54
Percent Similarity: 49.68%   Conservative: 23
Best Local Similarity: 34.84%   Mismatches: 49
Query Match: 15.30%   Indels: 29
DB: Gaps: 4

US-10-054-313-1 (1-286) x AAF94380 (1-462)

Qy   140 ValValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGly 159
             ::: ::::::||||||||| |||||| ||||||
Db   13 ATTGAAATTTTACTGATGGATCTTGCTTAGGTAAT--------------CCAGGGGCGGGC 60

Qy   160 IleGlyValTyrTrpGlyHisProLeuAsnValGlyIleArgLeuProGlyArg 179
             ::||||| || ::: |||
Db   61 ------------------------------GGAATTGGTGCCGTATTGCGTTATAAA 87

Qy   180 Gln------------------------------ThrAsnGlnArgAlaGluIle 187
             ::: |||:::||||||||||
Db   88 CAACATGAAAAAACACTCTCCAAAGGCTATTCCAAACCACCAATAATCGAATTA 147

Qy   188 HisAlaAlaCysLysAlaIleGluGlnAlaLysThrGlnAsnIleAsnLysLeuValLeu 207
             ||||||||||||||| :::
Db   148 CGGCGCTGTCATTGAAGCATTAAATACATTAAAAGAACCTTGCTTG------ATCACGCTT 201

Qy   208 TyrThrAspSerMetPheThrIleAsnGlyIleThrAsnTrpValGlnGlyTrpLysLys 227
             ||||||||| :::   ::: ::||||||||||||| ||||||
Db   202 TATAGTGATAGCCAATAATGAAAAAATGGCATAACCAAATGGATCTTTAACTGGAAAAAA 261

Qy   228 AsnGlyTrpLysThrSerAlaGlyLysGluValIleLeuAsnLysGluAspPheValAlaLeu 247
             ||||||||| :::||||||
Db   262 AATAATTGGAAAGCAAGTTCTGGAAAGCCTGTAAAAAACCAAGATTTATGGATAGCCTTA 321

Qy   248 GluArgLeuThrGlnGlyMetAspIleGlnTrpMetHisValProGlyHisSerGlyPhe 267
             ::: |||||||||||:::
Db   322 GATGAATCCATCCAACGTCATAAAAATTAATTGGCAATGGGTAAAAGGCCATGCTGGACAC 381

Qy   268 IleGlyAsnGluGlnAlaAspArgLeuAlaArgGluGlyValGlyAlaLys 282
             :::   ::||||| ||||
Db   382 AGAGAAAATGAAGATGAATTTGCGATGAATTAGCAAAAAAAGGGGCAGAA 426

RESULT 15
ABK64944
ID   ABK64944 standard; DNA; 462 BP.
XX
AC   ABK64944;
XX
DT   18-JUN-2002 (first entry)
XX
DE   DNA encoding Haemophilus influenzae essential gene #36.
XX
KW   Essential bacterial gene; antifungal agent; antibacterial agent;
KW   antiparasitic agent; insecticidal agent; microbial infection;

KW   mucous membrane infection; otitis media; sinusitis; bronchitis;
KW   alveolitis; conjunctivitis; pneumonia; meningitis; epiglottis;
KW   cellulitis; septic arthritis; gene; ds.
XX
OS   Haemophilus influenzae.
XX
PN   WO200218601-A2.
XX
PD   07-MAR-2002.
XX
PF   22-AUG-2001; 2001WO-US26245.
XX
PR   25-AUG-2000; 2000US-0649145.
XX
PA   (ABBO) ABBOTT LAB.
XX
PI   Chovan LE, Hessler PE,   Reich KA;
XX
DR   WPI; 2002-304258/34.
DR   P-PSDB; AAU91458.
XX
PT   Essential bacterial genes in Haemophilus influenzae necessary for
PT   bacterium's growth and survival, useful for screening inhibitors of
PT   polypeptides and developing therapeutic agents e.g. antimicrobial -
XX
PS   Claim 1; Page 116-117; 185pp; English.
XX
CC   The invention describes an essential bacterial gene (I) comprising a
CC   purified polynucleotide isolated from Haemophilus influenzae where (I)
CC   is essential to H. influenzae survival. The encoded polypeptide (II) is
CC   useful for screening substances that function to inhibit essential H.
CC   influenzae polypeptides by contacting (II) with the desired substances
CC   and measuring the response by a screen from specific, enzyme, general,
CC   affinity, phenotypic and binding screen. (I) and (II) are useful in
CC   developing therapeutic agents such as antifungal, antibacterial and
CC   antiparasitic agent, insecticidal agent, and preventive antimicrobial
CC   agents which are effective in preventing microbial infection or useful
CC   in treatment of that particular infection. (I) and (II) may also be
CC   useful in treatment of mucous membrane infections such as otitis media,
CC   sinusitis, bronchitis, alveolitis, conjunctivitis, pneumonia, meningitis,
CC   epiglottis, cellulitis and septic arthritis. This sequence encodes an
CC   essential H. influenzae gene, described in the invention.
XX
SQ   Sequence 462 BP; 180 A; 74 C; 96 G; 111 T; 1 other;

Alignment Scores:
Pred. No.:   3.25e-12   Length: 462
Score: 236.50   Matches: 54
Percent Similarity: 49.68%   Conservative: 23
Best Local Similarity: 34.84%   Mismatches: 49
Query Match: 15.30%   Indels: 29
DB: Gaps: 4

US-10-054-313-1 (1-286) x ABK64944 (1-462)

Qy   140 ValValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGly 159
             ::: ::::::||||||||| |||||| ||||||
Db   13 ATTGAAATTTTACTGATGGATCTTGCTTAGGTAAT--------------CCAGGGGCGGGC 60

Qy   160 IleGlyValTyrTrpGlyHisProLeuAsnValGlyIleArgLeuProGlyArg 179
             ::||||| || ::: |||
Db   61 ------------------------------GGAATTGGTGCCGTATTGCGTTATAAA 87

Qy   180 Gln------------------------------ThrAsnGlnArgAlaGluIle 187
             ::: |||:::||||||||||
Db   88 CAACATGAAAAAACACTCTCCAAAGGCTATTCCAAACCACCAATAATCGAATTA 147

Qy   188 HisAlaAlaCysLysAlaIleGluGlnAlaLysThrGlnAsnIleAsnLysLeuValLeu 207
             ||||||||||||||| :::
Db   148 CGGCGCTGTCATTGAAGCATTAAATACATTAAAAGAACCTTGCTTG------ATCACGCTT 201

Qy   208 TyrThrAspSerMetPheThrIleAsnGlyIleThrAsnTrpValGlnGlyTrpLysLys 227
             ||||||||| :::   ::: ::||||||||||||| :::

```
Db     202  TATAGTGATAGCCAATATATGAAAAATGGCATAACCAAATGGATCTTAACTGGAAAAA  261

Qy     228  AsnGlyTrpLysThrSerAlaGlyLyLysGluValIleAsnLysGluAspPheValAlaLeu  247
                ||||||    |||::||||           :|::::::|||||||
Db     262  AATAATTGGAAAGCAAGTTCTGGAAAGCCTGTAAAAAACCAAGATTTATGGATAGCCTTA  321

Qy     248  GluArgLeuThrGlnGlyMetAspIleGlnTrpMetHisValProGlyHisSerGlyPhe  267
               :::       |||                ||||||:::||
Db     322  GATGAATCCATCCAACGTCATAAAATTAATTGGCAATGGGTAAAAGGCCATGCTGGACAC  381

Qy     268  IleGlyAsnGluGluAlaAspArgLeuAlaArgGluGlyAlaLys  282
               |||||    |||:::::|||||||
Db     382  AGAGAAAATGAAATTTGCGATGAATTAGCAAAAAAAGGGGGCAGAA  426
```

Search completed: December 17, 2003, 19:23:34
Job time : 387 secs

```
                    OM protein - nucleic search, using frame_plus_p2n model

                         Copyright (c) 1993 - 2003 Compugen Ltd.

                              GenCore version 5.1.6

Run on:        December 17, 2003, 19:11:56 ; Search time 73 Seconds
                                            (without alignments)
                                            1729.254 Million cell updates/sec

Title:         US-10-054-313-1
Perfect score: 1546
Sequence:      1 MSWLLFLAHRVALAALPCRR.........FIGNEEADRLAREGAKQSED 286

Scoring table: BLOSUM62
               Xgapop 10.0 , Xgapext  0.5
               Ygapop 10.0 , Ygapext  0.5
               Fgapop  6.0 , Fgapext  7.0
               Delop   6.0 , Delext   7.0

Searched:      569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters:        1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries
```

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10054313/runat_17122003_150744_24207/app_query.fasta_1.455
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10054313 @CGN_1_156 @runat_17122003_150744_24207 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -XGAPOP=10
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :   Issued_Patents_NA:*
              1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
              2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
              3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
              4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
              5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
              6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                              SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	241	15.6	1275	4	US-09-252-991A-13638	Sequence 13638, A
2	236.5	15.3	1830121	4	US-09-557-884-1	Sequence 1, Appli
3	236.5	15.3	1830121	4	US-09-643-990A-1	Sequence 1, Appli
4	235.5	15.2	801	4	US-09-199-637A-154	Sequence 154, App
5	235.5	15.2	2048	4	US-09-199-637A-136	Sequence 136, App
c 6	233.5	15.1	570	4	US-09-252-991A-13802	Sequence 13802, A
7	217	14.0	1509	4	US-09-328-352-1765	Sequence 1765, Ap
8	174	11.3	501	1	US-08-163-181-3	Sequence 3, Appli
9	174	11.3	501	1	US-08-465-161-3	Sequence 3, Appli
10	174	11.3	552	1	US-08-163-181-4	Sequence 4, Appli
11	174	11.3	552	1	US-08-465-161-4	Sequence 4, Appli
12	160	10.3	8088	4	US-09-315-127-4	Sequence 4, Appli
13	160	10.3	8535	3	US-08-716-351A-1	Sequence 1, Appli
14	157	10.2	8060	3	US-08-766-528-1	Sequence 13, Appl
c 15	156	10.1	8202	1	US-08-258-420-13	Sequence 13, Appl
16	154	10.1	3612	2	US-09-265-013-3	Sequence 3, Appli
17	154	10.0	7308	3	US-09-011-745-3	Sequence 3, Appli
18	154	10.0	7308	3	US-09-011-745-4	Sequence 4, Appli
19	154	10.0	7616	3	US-09-011-745-2	Sequence 2, Appli
20	154	10.0	8332	3	US-08-850-961-1	Sequence 1, Appli
21	154	10.0	8332	4	US-09-479-776-1	Sequence 1, Appli
22	154	10.0	8332	4	US-09-309-572-11	Sequence 11, Appl
23	154	10.0	8332	4	US-09-315-127-1	Sequence 1, Appli
24	154	10.0	8332	4	US-09-265-013-1	Sequence 1, Appli
25	154	10.0	8332	4	US-09-554-572-25	Sequence 25, Appl
26	153	9.9	6363	2	US-08-929-967-6	Sequence 6, Appli
27	152	9.8	8323	2	US-08-110-300A-8	Sequence 8, Appli
28	152	9.8	8323	2	US-08-886-642-8	Sequence 8, Appli
29	152	9.8	8323	4	US-09-433-322B-2	Sequence 2, Appli
30	152	9.8	8323	4	PCT-US93-08041-8	Sequence 8, Appli
31	152	9.8	10367	1	US-08-110-300A-9	Sequence 9, Appli
c 32	152	9.8	10367	5	US-08-886-642-9	Sequence 9, Appli
c 33	152	9.8	10367	5	PCT-US99-08041-9	Sequence 9, Appli
c 34	149	9.6	8132	3	US-08-766-528-3	Sequence 3, Appli
35	144	9.3	8655	1	US-09-075-272-1	Sequence 9, Appli
36	140.5	9.1	9472	1	US-08-325-547-9	Sequence 9, Appli
37	139.5	9.0	9468	1	US-08-325-547-10	Sequence 10, Appl
38	135.5	8.8	3093	4	US-08-811-682-14	Sequence 14, Appl
39	135.5	8.8	3841	4	US-08-811-68-13	Sequence 13, Appl
40	135.5	8.8	9471	4	US-08-811-682-1	Sequence 1, Appli
41	132.5	8.6	9633	6	5223423-1	Patent No. 5223423
42	127.5	8.2	2689	2	US-08-876-546A-15	Sequence 15, Appl
43	127.5	8.2	2689	3	US-09-412-252-15	Sequence 15, Appl
44	127.5	8.2	3168	3	US-08-659-251-14	Sequence 14, Appl
45	127.5	8.2	3168	3	US-09-256-490-14	Sequence 14, Appl

                              ALIGNMENTS

RESULT 1
US-09-252-991A-13638
; Sequence 13638, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 13638
; LENGTH: 1275
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13638

Alignment Scores:
Pred. No.:   Length: 1275
Score: 3.79e-17   Matches: 92
Percent Similarity: 241.00   Conservative: 36
Best Local Similarity: 35.96%   Mismatches: 126
Query Match: 25.84%   Indels: 102
DB: 15.59%   Gaps: 14
                        4

US-10-054-313-1 (1-286) x US-09-252-991A-13638 (1-1275)

Qy   9 HisArgValAlaLeuAlaAlaLeuProCysArgArgGlySerArgGlyPheGlyMetPhe 28
          ||| ::: ||| ::|||| ||| :::
Db 259 CACCGGGCAGATTCAGCGGCG------TGTGCGCCTGGGGCCCCGCCGTGCCCGGCGTGGA 312
```

APPLICANT: Fleischmann et al.
TITLE OF INVENTION: The Nucleotide sequence of
           the Haemophilus influenzae Rd Genome, Fragments
           Thereof, and Uses Thereof

NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
   ADDRESSEE: Human Genome Sciences, Inc.
   STREET: 9410 Key West Avenue
   CITY: Rockville
   STATE: MD
   COUNTRY: USA
   ZIP: 20850
COMPUTER READABLE FORM:
   MEDIUM TYPE: 3 1/2 inch diskette
   COMPUTER: Dell Pentium
   OPERATING SYSTEM: MS DOS v6.22
   SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
   APPLICATION NUMBER: US/09/557,884
   FILING DATE: 25-Apr-2000
   CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
   APPLICATION NUMBER: 08/476,102
   FILING DATE: JUN-5-1995
ATTORNEY/AGENT INFORMATION:
   NAME: Michelle S. Marks
   REGISTRATION NUMBER: 41,971
   REFERENCE/DOCKET NUMBER: PB186P3
TELECOMMUNICATION INFORMATION:
   TELEPHONE: 301-309-8504
   TELEFAX: 301-309-8439
INFORMATION FOR SEQ ID NO: 1:
   SEQUENCE CHARACTERISTICS:
   LENGTH: 1830121 base pairs
   TYPE: nucleic acid
   STRANDEDNESS: double
   TOPOLOGY: linear
   SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1

Alignment Scores:
Pred. No.:   1830121
Score: 8.37e-12   Length: 1830121
Percent Similarity: 236.50   Matches: 54
Best Local Similarity: 49.68%   Conservative: 23
Query Match: 34.84%   Mismatches: 49
DB: 15.30%   Indels: 29
                        4   Gaps: 4

US-10-054-313-1 (1-286) x US-09-557-884-1 (1-1830121)

Qy   29 TyrAlaValArgGlyArgGlyThrGlyValPheLeuThrTrpAsnGluCysArgAla 48
          : :::||||||||| ||||
Db   313 CATCGGCCTGCGAAGAGGGCGC------------CTGCCGCTCAGCGAACATGCCGCGGA 360

Qy   49 GlnValAspArgPheProAlaAlaArg--------PheLysLysPheAlaThrGluAsp 65
                            ||| |||
Db   361 ----CGTTGGTCCTGCTGCAACACGGCCTGGATTTCTGCCTGTCGCCTCACCGTCT 411

Qy   66 GluAlaTrpAla---PheValArgLysSerAlaSerProGluValSerGluGlyHisGlu 84
          ||| ||| ::: ::::
Db   412 CCTGCGCGGAAGCCGCGGTACCGTTCGTTCGGGCGGCGGCCACCTGCTGCTGATCGGCATCAA 471

Qy   85 AsnGlnHis---------------------GlyGln 89
          |||
Db   472 CCCATGGAGCCGTTGGGGACGCATCGTCATTATTTCGCCGGGGGATGCCTTGCCCCAGGCCCG 531

Qy   90 GluSerGluAlaLysProGly-LysArgLeuArgGluProLeuAspGlyHisGluHisGlu 109
                :::
Db   532 CTGCATTCCTCCGTCGCCGGCCTGCCGACTGGCTCAACCTGCTGGGCTTCGCCCTGGAGAA 591

Qy   110 SerAla--- 111
          :::
Db   592 ACGGCGCTTCGGGTGCTATCGTCCGGCCGCTTGCGTCCGGCAGCCTGCAATCGCGCCTGGC 651

Qy   112 GlnProTyrAlaLysHisMetLys--------------------- 119
          :::
Db   652 TCGCCTGGAGCGGCTGGGGGCGCCGCTGCAGTCTTCGGGCGCGGCCTTCTATCTATTGGT 711

Qy   120 ----------ProSerValGluProAlaProValSerArgArgAspThr------------ 132
                        :::
Db   712 GGCACGCAAGCTGGTCGTGGGGTTGCGCCGTTGCGCCAGAGCAAGCGGCAACCGCGCGCGG 771

Qy   133 --------------------------PheSerTyrMetGly--- 137
                                        ||| :::
Db   772 TCAGCTGGTGCCCATGCCGGTGGCGAAAGTCAGCCGGCGGAGAATTCCGAAATTTAGGCATG 831

Qy   138 -----------AspPheValValTyrThrAspGlyCysCysSerSerAsnGlyArgArg 154
                       :::
Db   832 ACAGATAAAGAACAGAGACGTAGTGATCTATACCGACGGCGCCTGCAAGGGCAACCCTGGGCGC 891

Qy   155 LysProArgAlaGlyIleGlyValTyr----------------TrpGlyProGly 167
                                      :::|||||| ||||||| |
Db   892 GGCGGCTGGGGGGCGTTGCTCCTCTACAAGGGGCGCGAGCGAGCTTTGGGGC---GGC 948

Qy   168 HisProLeuAsnValGlyIleArgLeuProGlyGlnThrArgGlnThrArgAlaGluIle 187
                        ::: ||||||
Db   949 ---ACCGACTCGGATACGTGATGCCGGCATCACCGGTTGCCGGAACTGGAAGAAG 981

Qy   188 HisAlaAlaCysLysAlaIleGluGlnGlnAlaLysThrGlnAsnIleAsnLysLeuValLeu 207
          ::::|||||| :::|||||||
Db   982 ATGGCGGGCGGATCCAGGCGCTGGGCGCCGCACTCAAGCCGTTCCGTCCGATCCGTCTGATC--- 1038

Qy   208 TyrThrAspSerMetPheThrIleAsnGlyIleThrAsnTrpValGlnGlyTrpLysLys 227
                    :::
Db 1039 ---ACCGACTCGGATACGTGATGCCGGCATCACCGGTTGCCGGAACTGGAAGAAG 1095

Qy   228 AsnGlyTrpLysThrSerAlaGlyLysLysGluValAlaIleAsnLysGluValAlaLeu 247
                                      :::
Db 1096 CGCGGCTGGAAGACCGACCAAGCAGCCGGTCAAGAATGCCGACCTCTGGCAGGCCCTG 1155

Qy   248 GluArgLeuThrGlnGlyMetAspIleGlnTrpMetHisValProGlyHisSerGlyPhe 267
                    :::
Db 1156 GATGAGCAGGCAGGTCGCCCGGCACCAGGTGGCAGTGGGTCCGGGCCATACCGGCGGCAT 1215

Qy   268 IleGlyAsnGluGluAlaAspArgLeuAlaAspArgGluGlyLeuAlaArgGluValAlaLysGln 283
                        :::
Db 1216 CCCGGCAACGAACAACAGCCCGACCAGTTGGCCAACCTGTGGCCTGGCGGTCGCCGAA 1263

Qy   140 ValValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGly 159
          : :::||||||
Db 152742 ATTGAAATTTTACTGATGGATCTTGCTTAGGTAAT------------CCAGGGGCGCGGC 152789

Qy   160 IleGlyIleValTyrTrpGlyHisProLeuAsnValGlyIleArgLeuProGlyArg 179
                                :::
Db 152790 ------------------------------GGAATTGGTGCCGTATTGGCGTTATAAA 152816

Qy   180 Gln----------------------ThrAsnGlnArgAlaGluIle 187
          |||
Db 152817 CAACATGAAAAAACACTCTCCAAAGGCTATTTCCAAACCACCAATAATCGAATGGAATTA 152876

Qy   188 HisAlaAlaCysLysAlaIleGluGlnGlnAlaLysThrGlnAsnIleAsnLysLeuValLeu 207
                                :::
Db 152877 CGCGGCTGTCATTGAAGCATTAAATACATTAAAAGAACCTTGCTTG-----ATCACGCTT 152930

Qy   208 TyrThrAspSerMetPheThrIleAsnGlyIleThrAsnTrpValGlnGlyTrpLysLys 227
                :::
Db 152931 TATAGTGATAGCCAATATATGAAAAATGCATAACCAAATGACTTTAACTGGAAAAA 152990

Qy   228 AsnGlyTrpLysThrSerAlaGlyLysLysGluValAlaIleAsnLysGluValAlaLeu 247
          :::
Db 152991 AATAATTGAAAGCAAGTTCTGAAAGCCTGTAAAAACCAAGATTTATGGATAGCCTTA 153050

RESULT 2
US-09-557-884-1
; Sequence 1, Application US/09557884
; Patent No. 6506581
; GENERAL INFORMATION:

```
Qy     248  GluArgLeuThrGlnGlyMetAspIleGlnTrpMetHisValProGlyHisSerGlyPhe  267
                ::|
Db  153051  GATGAATCCATCCAACGTCATAAAATTAATTGGCAATGGTAAAAGGCCATGCTGGACAC  153110

Qy     268  IleGlyAsnGluGluAlaAspArgLeuAlaArgGluGluGlyAlaGlyLys  282
            ||||||::::::|||||
Db  153111  AGAGAAATGAATTGCGATGAATTAGCAAAAAAAGGGGCCAGAA  153155

RESULT 3
US-09-643-990A-1
; Sequence 1, Application US/09643990A
; Patent No. 6528289
; GENERAL INFORMATION:
;   APPLICANT: Robert D. Fleischmann
;              Mark D. Adams
;              Owen White
;              Hamilton O. Smith
;              J. Craig Venter
;   TITLE OF INVENTION: The Nucleotide sequence of
;              the Haemophilus influenzae Rd Genome, Fragments
;              Thereof, and Uses Thereof
;   NUMBER OF SEQUENCES:1
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Human Genome Sciences, Inc.
;     STREET: 9410 Key West Avenue
;     CITY: Rockville,
;     STATE: MD
;     COUNTRY: USA
;     ZIP: 20850
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: 3 1/2 inch diskette
;     COMPUTER: Dell Pentium
;     OPERATING SYSTEM: MS DOS v6.22
;     SOFTWARE: ASCII Text
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/09/643,990A
;     FILING DATE: 23-Aug-2000
;     CLASSIFICATION: <Unknown>
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: 08/487,429
;     FILING DATE: 1995-06-07
;     APPLICATION NUMBER: 08/426,787
;     FILING DATE: 1995-04-21
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Kenley K. Hoover
;     REGISTRATION NUMBER: 40,302
;     REFERENCE/DOCKET NUMBER: PB186P1C1
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 301-610-5790
;     TELEFAX: 310-309-8439
; INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 1830121 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: double
;     TOPOLOGY: linear
;   SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1

Alignment Scores:
Pred. No.:               8.37e-12        Length:        1830121
Score:                   236.50          Matches:       54
Percent Similarity:      49.68%          Conservative:  23
Best Local Similarity:   34.84%          Mismatches:    49
Query Match:             15.30%          Indels:        29
DB:                      4               Gaps:          4

US-10-054-313-1 (1-286) x US-09-643-990A-1 (1-1830121)

Qy     140  ValValValTyrThrAspGlyCysCysSerSerAsnGlyLysArgArgLysProArgAlaGly  159
                :::   :::|||||||||||||
Db  152742  ATTGAAATTTTACTGATGGATCTTGCTTAGGTAAT----------CCAGGGGCCGGGC  152789
```

```
Qy     160  IleGlyValTyrTrpProGlyHisProLeuAsnValGlyIleArgLeuProGlyArg  179
                                                  :::|||
Db  152790  --------------------------------------GGAATTGGTGCCGTATTGCGTTATAAA  152816

Qy     180  Gln------------------------------ThrAsnGlnArgAlaGluIle  187
            ||:                              |||||:::
Db  152817  CAACATGAAAAAACACTCTCCAAAGGCTATTCCAAACCAATAATCGAATGAATGGAATTA  152876

Qy     188  HisAlaAlaCysLysAlaIleGluGlnAlaLysThrGlnAsnIleAsnLysLeuValLeu  207
                :::|||| :::|:|
Db  152877  CGGCGCTGTCATTGAAGCATAAATACATTAAAAGAACCTTGCTTG-----ATCACGCTT  152930

Qy     208  TyrThrAspSerMetPheThrIleAsnGlyIleThrAsnTrpValGlnGlyTrpLysLys  227
            ||:|||||    |||:||
Db  152931  TATAGTGATAGCCAATATATGAAAAATGGCATAACCAAATGGATCTTTAACTGGAAAAAA  152990

Qy     228  AsnGlyTrpLysThrSerAlaIleGlyLysGluValIleGluAsnLysGluAspPheValAlaLeu  247
                :::||| ::::||||||
Db  152991  AATAATTGGAAAGCAAGTTCTGGAAAGCCTGTAAAAAACCAAGATTTATGATAGCCTTA  153050

Qy     248  GluArgLeuThrGlnGlyMetAspIleGlnTrpMetHisValProGlyHisSerGlyPhe  267
                :::
Db  153051  GATGAATCCATCCAACGTCATAAAATTAATTGGCAATGGTAAAAGGCCATGCTGGACAC  153110

Qy     268  IleGlyAsnGluGluAlaAspArgLeuAlaArgGluGluGlyAlaGlyLys  282
            ||||||:::::|||||
Db  153111  AGAGAAATGAATTGCGATGAATTAGCAAAAAAAGGGGCCAGAA  153155

RESULT 4
US-09-199-637A-154
; Sequence 154, Application US/09199637A
; Patent No. 6355411
; GENERAL INFORMATION:
;   APPLICANT: Ausubel, Frederick
;   APPLICANT: Goodman, Howard M.
;   APPLICANT: Rahme, Laurence G.
;   APPLICANT: Mahajan-Miklos, Shalina
;   APPLICANT: Tan, Man-Wah
;   APPLICANT: Cao, Hui
;   APPLICANT: Drenkard, Eliana
;   APPLICANT: Tsongalis, John
;   TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
;   TITLE OF INVENTION: SEQUENCES AND USES THEREOF
;   FILE REFERENCE: 00786/361002
;   CURRENT APPLICATION NUMBER: US/09/199,637A
;   CURRENT FILING DATE: 1998-11-25
;   PRIOR APPLICATION NUMBER: 60/066,517
;   PRIOR FILING DATE: 1997-11-25
;   NUMBER OF SEQ ID NOS: 437
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 154
;     LENGTH: 801
;     TYPE: DNA
;     ORGANISM: Pseudomonas aeruginosa
US-09-199-637A-154

Alignment Scores:
Pred. No.:               7.55e-17        Length:        801
Score:                   235.50          Matches:       60
Percent Similarity:      49.12%          Conservative:  24
Best Local Similarity:   35.09%          Mismatches:    64
Query Match:             15.23%          Indels:        24
DB:                      4               Gaps:          5

US-10-054-313-1 (1-286) x US-09-199-637A-154 (1-801)

Qy     124  ProAlaProValSerArgAspThrPheSerTyrMetGly--------------AspPhe  139
                                      :::|||
Db     314  CCGTGGCGAAAGTCAGCCGGCG-AGATTCCGAAATTAGGCATGACAGATAAAGAACAG  372

Qy     140  ValValValTyrThrArgAspGlyCysCysSerSerAsnGlyLysArgArgLysProArgAlaGly  159
            |||||||
```

```
Db  589  GTAGTGATCTATACCGACGGCGCCTGCAAGGGCCAACCCTGGGCGGCGGGCGG  530
              |||||||::::|||||||||||||  |||
Qy  160  IleGlyValTyr------------TrpGlyHisProGlyLeuAsnVal  172
              :::  :::|||
Db  529  TTGCTCCTCTACAAGGGCGCCGAGCGAGAGCTTTGGGGC---GGCGAGCCG  482
              ||||     |||
Qy  173  GlyIleArgLeuProGlyArgGlnThrArgGlnArgIleHisAlaAlaCysLys  192
              ||||||  ::|
Db  481  ---------GACACCACCAACCGCATGGAGCTGATGGCGGCGATCCAG  440
Qy  193  AlaIleGluGlnAlaLysThrGlnAsnIleAsnIleAsnLysAsnLysValTyrThrAspSerMet  212
              ||::  ||  ::::|||||::|
Db  439  GCGCTGGCGGCCACTCAAGGCGTTCCTGTCCGATCCGTCTGATC------ACCGACTCGGAA  386
Qy  213  PheThrIleAsnGlyIleThrAsnTrpProValGlnGlyTrpLysAsnGlyTyrTrpLysThr  232
              :::  :::|||:::|
Db  385  TACGTGATGCGCGGCATCACCGAATGGTTGCCGAACTGGAGAAGAGCCGGGCTGGAAGACC  326
Qy  233  SerAlaGlyLysGluValIleIleAsnLysGluAsnLysAspPheValAlaLeuGluArgLeuThrGln  252
              ::::   :::  :::|
Db  325  GCCAAGCAGCAGCCTGTCAAGAATGCCGACCTCTGGCAGCCCCTGGATGAACAGGTCGCC  266
Qy  253  GlyMetAspIleGlnTrpMetHisValProGlyHisSerGlyPheIleGlyAsnGluGlu  272
              ::::   :::||:::|
Db  265  CGGCACCAGGTGGAGTGGCAGTGGCGTCCGGGCATACCGGGCACCCCGGCAACGAGCGG  206
Qy  273  AlaAspArgLeuAlaArgGluGlyAlaLysGln  283
              ||||:::|||||
Db  205  GCCGACCAGTTGGCCAACCGTGGCGTCGCGCCGAA  173
```

RESULT 6
US-09-252-991A-13802/c
; Sequence 13802, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 13802
; LENGTH: 570
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13802

Alignment Scores:

Pred. No.:	Length:	570
Score: 7.46e-17	Matches:	54
233.50	Conservative:	22
Percent Similarity: 50.33%	Mismatches:	56
Best Local Similarity: 35.76%	Indels:	19
Query Match: 15.10%	Gaps:	4
DB: 4		

US-10-054-313-1 (1-286) x US-09-252-991A-13802 (1-570)

```
Qy  140  ValValValTyr------------TrpGlyHisProGlyLeuAsnVal  172
              |||||||::::|||||||||||||
Db  546  GTAGTGATCTATACCGACGGCGCCTGCAAGGGCCAACCCTGGGCGGCGGGCGG  487
Qy  160  IleGlyValTyr----
              :::  :::|||
Db  486  TTGCTCCTCTACAAGGGCGCCGAGCGAGAGCTTTGGGGC---GGCGAGCCG  439
Qy  173  GlyIleArgLeuProGlyArgGlnThrArgGlnArgIleHisAlaAlaCysLys  192
              ||||||  ::|
```

```
Db  373  GTAGTGATCTATACCGACGGCGCCTGCAAGGGCGCAACCCTGGGCGGCGGGCGG  432
              |||||||::::||||||||||||||
Qy  160  IleGlyValTyr------------TrpGlyProGlyHisProLeuAsnVal  172
              ::  :::|||
Db  433  TTGCTCCTCTACAAGGGCGCCGAGCGAGAGCTTTGGGGC---GGCGAGCCG  480
              ||||  |||
Qy  173  GlyIleArgLeuProGlyArgGlnThrArgAsnGlnArgAlaGluHisAlaAlaCysLys  192
              ||||  :::|
Db  481  ---------GACACCACCAACCGCATGGAGCTGATGGCGGCGATCCAG  522
Qy  193  AlaIleGluGlnAlaLysThrGlnAsnIleAsnIleAsnLysAsnLysValThrThrAspSerMet  212
              ||:::  ||  ::::|
Db  523  GCGCTGGCGGCCACTCAAGGCGTTCCTGTCCGATCCGTCTGATC------ACCGACTCGGAA  576
Qy  213  PheThrIleAsnGlyIleThrAsnTrpProValGlnGlyTrpLysAsnGlyTyrTrpLysThr  232
              :::  :::|||:::|
Db  577  TACGTGATGCGCGGCATCACCGAATGGTTGCCGAACTGGAGAAGAGCCGGGCTGGAAGACC  636
Qy  233  SerAlaGlyLysGluValIleIleAsnLysGluAsnLysAspPheValAlaLeuGluArgLeuThrGln  252
              ::::   :::  :::|
Db  637  GCCAAGCAGCAGCCTGTCAAGAATGCCGACCTCTGGCAGCCCCTGGATGAACAGGTCGCC  696
Qy  253  GlyMetAspIleGlnTrpMetHisValProGlyHisSerGlyPheIleGlyAsnGluGlu  272
              ::::   :::||:::|
Db  697  CGGCACCAGGTGGAGTGGCAGTGGCGTCCGGGCATACCGGGCACCCCGGCAACGAGCGG  756
Qy  273  AlaAspArgLeuAlaArgGluGlyAlaLysGln  283
              ||||:::|||
Db  757  GCCGACCAGTTGGCCAACCGTGGCGTCGCGCCGAA  789
```

RESULT 5
US-09-199-637A-136/c
; Sequence 136, Application US/09199637A
; Patent No. 6355411
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick
; APPLICANT: Goodman, Howard M.
; APPLICANT: Rahme, Laurence G.
; APPLICANT: Mahajan-Miklos, Shalina
; APPLICANT: Tan, Man-Wah
; APPLICANT: Cao, Hui
; APPLICANT: Drenkard, Eliana
; APPLICANT: Tsongalis, John
; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF
; FILE REFERENCE: 00786/361002
; CURRENT APPLICATION NUMBER: US/09/199,637A
; CURRENT FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: 60/066,517
; PRIOR FILING DATE: 1997-11-25
; NUMBER OF SEQ ID NOS: 437
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 136
; LENGTH: 2048
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-199-637A-136

Alignment Scores:

Pred. No.:	Length:	2048
Score: 3.19e-16	Matches:	60
235.50	Conservative:	24
Percent Similarity: 49.12%	Mismatches:	64
Best Local Similarity: 35.09%	Indels:	24
Query Match: 15.23%	Gaps:	5
DB: 4		

US-09-199-637A-136 (1-286) x US-09-199-637A-136 (1-2048)

```
Qy  124  ProAlaProProValSerArgArgAspThrPheSerTyrMetGly--------AspPhe  139
              |||||  :::
Db  648  CCGGTGCGAAAGTCAGCCGCCG-AGATTCCGAAATTAGGCATGACAGATAAAGAACAG  590
              ||||  :::
Qy  140  ValValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGly  159
```

```
Db   438  --------------GACACCACCAAACAACCGCATGGAACTGATGGGCGGATCCAG  397

Qy   193  AlaIleGluGlnAlaLysThrGlnAsnIleAsnLysLeuValLeuTyrThrAspSerMet  212
               ::||||||||         :||||||
Db   396  GGGCTGGCGGCCACTCAAGCGTTCCTGTCCGATCCGTCTGATC------ACCGACTCGGAA  343

Qy   213  PheThrIleAsnGlyIleThrAsnTrpValGlnGlyIleTrpLysLysAsnGlyTrpLysThr  232
              ::|||   ||||||||          ||||||
Db   342  TACGTGATGCGCGCCATCACCGAATGGTTGCCGAACTGGAAGAAGCGCGGCTGGAAGACC  283

Qy   233  SerAlaGlyLysGluValIleAsnLysGluAspPheValAlaLeuGlnArgLeuThrGln  252
             :::|||||  ||||||       ::  |||||||
Db   282  GCCAGCAAGCAGCCGGTCAAGAATGCCGACCTCTGGCAGGCCCTGGATGAGCAGGTCGCC  223

Qy   253  GlyMetAspIleGlnTrpMetHisValProGlyHisSerGlyPheIleGlyAsnGluGlu  272
              ::::|||||||        |||    :::|||
Db   222  CGGCACCAGGTGCAGTGGCAGTGGCTCCGGGCATCCCGGGCCAACGAGCGCGG  163

Qy   273  AlaAspArgLeuAlaArgGluGlyGlyAlaLysGln  283
              ||||||  |||   :::     ||
Db   162  GCCGACACAGGTTGGCCAACCGTGGCGTCGCCGAA  130
```

```
RESULT 7
US-09-328-352-1765
; Sequence 1765 Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
;  APPLICANT: Gary L. Breton et al.
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
;  TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC99-03PA
;  CURRENT APPLICATION NUMBER: US/09/328,352
;  CURRENT FILING DATE: 1999-06-04
;  NUMBER OF SEQ ID NOS: 8252
;  SEQ ID NO 1765
;  LENGTH: 1509
;  TYPE: DNA
;  ORGANISM: Acinetobacter baumannii
US-09-328-352-1765
```

```
Alignment Scores:
Pred. No.:              2.24e-14        Length:        1509
Score:                  217.00          Matches:       52
Percent Similarity:     46.06%          Conservative:  24
Best Local Similarity:  31.52%          Mismatches:    53
Query Match:            14.04%          Indels:        36
DB:                     4               Gaps:          6

US-10-054-313-1 (1-286) x US-09-328-352-1765 (1-1509)
```

```
Qy   133  PheSerTyrMetGlyAspPheValValValTyrThrAspGlyCysCysSerSerAsnGly  152
             :::|||||       ||   :::|||||||||||||
Db   121  TTTCGATTTATGTCTCAAACAATCACACTGTATGTTGATGGTGCATGTCGTGCAAT---  177

Qy   153  ArgArgLysProArgAlaGlyAlaGlyIleGlyValTyrValTyrTrpGlyPro-------  166
              ||||||||               ||||||
Db   178  ---------------CCCGGGTTTAGGC---GGTTGGGGTGCATATGTCATTACCGAGCAG  219

Qy   167  ------------------GlyHisProLeuAsnValGlyIleArgLeuProGlyArg  179
                                :::|||||||
Db   220  GGTGAACATAAAATTATTTGGCGGTGAACCC---------------------------GAC  252

Qy   180  GlnThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaIleGluGlnAlaLysThr  199
              |||    :::    ||||    :::
Db   253  ACAACAAATAACCGTATGGAGCTTACTGCGGCCTATTGAAGGAATTTCTTTTTGCCCTCCT  312

Qy   200  GlnAsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThr  219
             :::|||         ||||||   ||||||
Db   313  GAT------GCTCAATTAAATTGTCTGGACAGACTCAAATTAATGTGAAGCAAGGTATTACA  366

Qy   220  AsnTrpValGlnGlyIleTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGluValIle  239
              :::|||||||    ||||||    :::|||
Db   367  GAATGGATTCAGGGTATTTGGAAAAAGAAAAACTGG-----------AAAGATGTTTAAA  411
```

```
Qy   240  AsnLysGluAspPheValAlaLeuAlaLeuGlnArgLeuThrGlnGlyMetAspIleGlnTrpMet  259
             |||||             ||||   :::      :::
Db   412  AATCCTGACCTATGGCAAAAACTGATGCCGTTTGTGCGGTAGAAATATTGAATGGAAC  471

Qy   260  HisValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgGlu  279
             :::  |||||||||:::|||
Db   472  TGGATTAAAGGCCATGCAGGACCATGCAGGTAATGAAATGCAGACCAGCTGCGAATTTA  531

Qy   280  GlyAlaLysGlnSer  284
             |||||||||
Db   532  GGTGCAGACAAAACT  546
```

```
RESULT 8
US-08-163-181-3
; Sequence 3, Application US/08163181
; Patent No. 5459055
; GENERAL INFORMATION:
;  APPLICANT: Jendrisak, Jerome J.
;  APPLICANT: Smith, Robert E.
;  APPLICANT: Dahl, Gary A.
;  TITLE OF INVENTION: THERMOSTABLE RIBONUCLEASE H
;  TITLE OF INVENTION: AND GENETIC CONSTRUCT
;  TITLE OF INVENTION: THEREFORE
;  NUMBER OF SEQUENCES: 4
;  CORRESPONDENCE ADDRESS:
;  ADDRESSEE: Quarles & Brady
;  STREET: 411 East Wisconsin Avenue
;  CITY: Milwaukee
;  STATE: Wisconsin
;  COUNTRY: U.S.A.
;  ZIP: 53202
;  COMPUTER READABLE FORM:
;  MEDIUM TYPE: Floppy disk
;  COMPUTER: IBM PC compatible
;  OPERATING SYSTEM: PC-DOS/MS-DOS
;  SOFTWARE: PatentIn Release #1.0, Version
;  SOFTWARE: #1.25
;  CURRENT APPLICATION DATA:
;  APPLICATION NUMBER: US/08/163,181
;  FILING DATE:
;  CLASSIFICATION: 435
;  PRIOR APPLICATION DATA:
;  APPLICATION NUMBER: US 07/815,095
;  FILING DATE: 27-DEC-1991
;  ATTORNEY/AGENT INFORMATION:
;  NAME: Baker, Jean C.
;  REGISTRATION NUMBER: P-35,433
;  REFERENCE/DOCKET NUMBER: 31-307-9001-1
;  TELECOMMUNICATION INFORMATION:
;  TELEPHONE: (414) 277-5709
;  TELEFAX: (414) 277-5774
; INFORMATION FOR SEQ ID NO: 3:
;  SEQUENCE CHARACTERISTICS:
;  LENGTH: 501 base pairs
;  TYPE: nucleic acid
;  STRANDEDNESS: double
;  TOPOLOGY: linear
;  MOLECULE TYPE: DNA (genomic)
US-08-163-181-3
```

```
Alignment Scores:
Pred. No.:              2.41e-10        Length:        501
Score:                  174.00          Matches:       51
Percent Similarity:     51.35%          Conservative:  25
Best Local Similarity:  34.46%          Mismatches:    59
Query Match:            11.25%          Indels:        14
DB:                     1               Gaps:          5

US-10-054-313-1 (1-286) x US-08-163-181-3 (1-501)

Qy   140  ValValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGly  159
             :::||||||||||
```

; MOLECULE TYPE: DNA (genomic)
US-08-465-161-3

Alignment Scores:
Pred. No.:
Score: 2.41e-10
Percent Similarity: 174.00
Best Local Similarity: 51.35%
Query Match: 34.46%
DB: 11.25%

Length: 501
Matches: 51
Conservative: 25
Mismatches: 59
Indels: 14
Gaps: 5

US-10-054-313-1 (1-286) x US-08-465-161-3 (1-501)

```
Qy  140 ValValValTyrThrAspGlyCysSerSerAsnGlyArgArgLysProArgAlaGly 159
            :::::|||:::::          ::                :::
Db   25 GTGGCCCTCTTCACCACGGGGCCTGCCTGGGAAACCC-CGGGCCCGG--------GGG 74

Qy  160 IleGlyValTyrTrpGlyPro--------GlyHisProLeuAsnValGlyIleArgLeu 176
            :::|||||                             :::||
Db   75 GTGGGC--------GGCCTTCCTCCGCTTCCACGCGCCCACGAGAAGCTCCTCCGGGGG 125

Qy  177 ProGlyArgGln-ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaIleGluGl 196
            ||||          :::  |||     ::: |||||
Db  126 AGAGGGCCTGCACCACCAACAACCGCATGGAGCTCAAGGCGGCCATAGAGGGCCTAAAGGC 185

Qy  196 nAlaLysThrGlnAsnIleAsnLysLeuAsnLeuValLeuTyrThrAspSerMetPheThrIleAs 216
            :::::                             |||||  :::
Db  186 CCTCAAGGAGCCT-------TGCGAGGTGGACCTCTACACCGACCAGCCACTACCTCCAAGAA 239

Qy  216 nGlyIleThr---AsnTrpValGlnGlyTrpTrpLysLysAsnGlyTrpLysThrSerAlaGl 235
            :::|||                 ::::||||:::||
Db  240 GGCCTTCACCGAGGGCTGGCTGGAAGGCTGGCGGGAAAAAGGGGCTGGCGGACGGGCGGAGGG 299

Qy  235 yLysGluValIleAsnLysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAs 255
            ::::       ::::::
Db  300 CAAGCCCGTGAAAAACCGACCTCTGGGAGCCCTCCTCTCGCCATGGCCCCCACCG 359

Qy  255 pIleGlnTrpMetHisValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspAr 275
            ||||||:::::||         :::::
Db  360 GGTGGCGTTCCACTTCGTGAAGGGGCACACGGGCACCCGGAGAACGAACGGGTGACCG 419

Qy  275 gLeuAlaArgGluGlyAlaLys 282
            ||||:::
Db  420 GGAGGGCGAGGCGCCAGGCCCAG 441
```

RESULT 10
US-08-163-181-4
; Sequence 4, Application US/08163181
; Patent No. 5459055
GENERAL INFORMATION:
  APPLICANT: Jendrisak, Jerome J.
  APPLICANT: Smith, Robert E.
  APPLICANT: Dahl, Gary A.
  TITLE OF INVENTION: THERMOSTABLE RIBONUCLEASE H
  TITLE OF INVENTION: AND GENETIC CONSTRUCT
  TITLE OF INVENTION: THEREFORE
  NUMBER OF SEQUENCES: 4
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Quarles & Brady
    STREET: 411 East Wisconsin Avenue
    CITY: Milwaukee
    STATE: Wisconsin
    COUNTRY: U.S.A.
    ZIP: 53202
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: PatentIn Release #1.0, Version
    SOFTWARE: #1.25
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/163,181
    FILING DATE:

```
Db       GTGGCCCTCTTCACCACGGGGCCTGCCTGGGAAACCC-CGGGCCCGG--------GGG 74

Qy  160 IleGlyValTyrTrpGlyPro--------GlyHisProLeuAsnValGlyIleArgLeu 176
            :::|||||                             :::||
Db   75 GTGGGC--------GGCCTTCCTCCGCTTCCACGCGCCCACGAGAAGCTCCTCCGGGGG 125

Qy  177 ProGlyArgGln-ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaIleGluGl 196
            ||||          :::  |||     ::: |||||
Db  126 AGAGGGCCTGCACCACCAACAACCGCATGGAGCTCAAGGCGGCCATAGAGGGCCTAAAGGC 185

Qy  196 nAlaLysThrGlnAsnIleAsnLysLeuAsnLeuValLeuTyrThrAspSerMetPheThrIleAs 216
            :::::                             |||||  :::
Db  186 CCTCAAGGAGCCT-------TGCGAGGTGGACCTCTACACCGACCAGCCACTACCTCCAAGAA 239

Qy  216 nGlyIleThr---AsnTrpValGlnGlyTrpTrpLysLysAsnGlyTrpLysThrSerAlaGl 235
            :::|||                 ::::||||:::||
Db  240 GGCCTTCACCGAGGGCTGGCTGGAAGGCTGGCGGGAAAAAGGGGCTGGCGGACGGGCGGAGGG 299

Qy  235 yLysGluValIleAsnLysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAs 255
            ::::       ::::::
Db  300 CAAGCCCGTGAAAAACCGACCTCTGGGAGCCCTCCTCTCGCCATGGCCCCCACCG 359

Qy  255 pIleGlnTrpMetHisValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspAr 275
            ||||||:::::||         :::::
Db  360 GGTGGCGTTCCACTTCGTGAAGGGGCACACGGGCACCCGGAGAACGAACGGGTGACCG 419

Qy  275 gLeuAlaArgGluGlyAlaLys 282
            ||||:::
Db  420 GGAGGGCGAGGCGCCAGGCCCAG 441
```

RESULT 9
US-08-465-161-3
; Sequence 3, Application US/08465161
; Patent No. 5500370
GENERAL INFORMATION:
  APPLICANT: Jendrisak, Jerome J.
  APPLICANT: Smith, Robert E.
  APPLICANT: Dahl, Gary A.
  TITLE OF INVENTION: THERMOSTABLE RIBONUCLEASE H AND GENETIC CONSTRUCTS THEREFORE
  NUMBER OF SEQUENCES: 4
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Quarles & Brady
    STREET: 411 East Wisconsin Avenue
    CITY: Milwaukee
    STATE: Wisconsin
    COUNTRY: U.S.A.
    ZIP: 53202
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: PatentIn Release #1.0, Version
    SOFTWARE: #1.25
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/465,161
    FILING DATE:
    CLASSIFICATION: 435
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 07/815,095
    FILING DATE: 27-DEC-1991
  ATTORNEY/AGENT INFORMATION:
    NAME: Baker, Jean C.
    REGISTRATION NUMBER: P-35,433
    REFERENCE/DOCKET NUMBER: 31-307-9001-1
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: (414) 277-5709
    TELEFAX: (414) 277-5774
INFORMATION FOR SEQ ID NO: 3:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 501 base pairs
    TYPE: nucleic acid
    STRANDEDNESS: double
    TOPOLOGY: linear
```

```
; ADDRESSEE: Quarles & Brady
; STREET: 411 East Wisconsin Avenue
; CITY: Milwaukee
; STATE: Wisconsin
; COUNTRY: U.S.A.
; ZIP: 53202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version
; SOFTWARE: #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465,161
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/815,095
; FILING DATE: 27-DEC-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Baker, Jean C.
; REGISTRATION NUMBER: P-35,433
; REFERENCE/DOCKET NUMBER: 31-307-9001-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (414) 277-5709
; TELEFAX: (414) 277-5774
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 552 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Other Nucleic Acid
Us-08-465-161-4

Alignment Scores:
Pred. No.:
Score:                  2.79e-10         Length:       552
Percent Similarity:     51.35%           Matches:      51
Best Local Similarity:  34.46%           Conservative: 25
Query Match:            11.25%           Mismatches:   59
DB:                                      Indels:       14
                                         Gaps:         5

US-10-054-313-1 (1-286) x US-08-465-161-4 (1-552)

Qy  140 ValValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGly 159
            ::|||    :|:|:||||||    :|
Db   76 GTGGCCCTTCTTCACCGACGGGGGCCTGCCTGGGAAACCC-CGGGCCCGG--------GGG 125

Qy  160 IleGlyValTyrTrpGlyPro-------GlyHisProLeuAsnValGlyIleArgLeu 176
            :|||                                            ::|||
Db  126 GTGGGC--------GGCCCTTCCCGCTTCCACGCCCCACGAGAAGTCTCTTCCGGGGG 176

Qy  177 ProGlyArgGln-ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaIleGluGl 196
            ||||||       :|:|    ::  ::|:|
Db  177 AGAGGCCTGCACCACCAACACCGCATGGAGGCGGCCCATAGAGGGGCCTAAAGGC 236

Qy  196 nAlaLysThrGlnAsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAs 216
                :|                                              :::
Db  237 CCTCAAGGAGGCCT------TGCGAGGTGGACCTCTACACCGACAGCCCACTACCTCAAGAA 290

Qy  216 nGlyIleThr--AsnTrpValGlnGlyTrpLysAsnGlyTrpLysThrSerAlaGl 235
            ||||        :|:|    ::|:||||||    ||||||:|
Db  291 GGCCTTCACCGAGGGCTGGCTGGAAGGCTGGCGGAAAAGGGGCTGGCGGCGGCGGAGG 350

Qy  235 YLysGluValIleAsnLysGluAspPheValAlaLeuGluThrGlnGlyMetAs 255
            |||         :::    ::   ||:|
Db  351 CAAGCCCGTGAAAAACCGGACCTCTGGGAGGCCCTCCTTCCTCCGCCATGGCCCCCACCG 410

Qy  255 pIleGlnTrpMetHisValProGlyHisSerGlyPheIleGlyAsnGluAlaAspAr 275
            :::::::::|:::||    ||||    :||
Db  411 GGTGCGCTTCCACTTCGTGAAGGGCACACGGCCACCCGGAGAACGAACGGGTGACCG 470
```

```
RESULT 11
US-08-465-161-4
; Sequence 4, Application US/08465161
; Patent No. 5500370
; GENERAL INFORMATION:
; APPLICANT: Jendrisak, Jerome J.
; APPLICANT: Smith, Robert E.
; APPLICANT: Dahl, Gary A.
; TITLE OF INVENTION: THERMOSTABLE RIBONUCLEASE H AND GENETIC CONSTRUCTS THEREFORE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
```

Qy   275 gLeuAlaArgGluGlyAlaLys 282
           ||||:||
Db   471 GGAGGCGAGGCGCCAGGCCCAG 492

RESULT 12
US-09-315-127-4
; Sequence 4, Application US/09315127
; Patent No. 6448390
; GENERAL INFORMATION:
; APPLICANT: The University of Tennessee, c/o Richard Cox
; TITLE OF INVENTION: Stable Envelope Proteins for Retroviral, Viral and
; TITLE OF INVENTION: Liposome Vectors and Use in Gene and Drug Therapy
; FILE REFERENCE: 44137-5023, U. of Tennessee
; CURRENT APPLICATION NUMBER: US/09/315,127
; CURRENT FILING DATE: 1999-05-20
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.0
;
; SEQ ID NO 4
; LENGTH: 8088
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:SEQ. ID NO. 3,
; OTHER INFORMATION: retroviral vector
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (5552)..(7552)
US-09-315-127-4

Alignment Scores:
Pred. No.:   Length: 8088
Score: 6.13e-07   Matches: 62
Percent Similarity: 43.00%   Conservative: 27
Best Local Similarity: 29.95%   Mismatches: 70
Query Match: 10.35%   Indels: 48
DB: 4   Gaps: 10

US-10-054-313-1 (1-286) x US-09-315-127-4 (1-8088)

Qy   90 GluSerGluAlaLysProGlyLysArgLeuArgGluProLeuAspGlyAspGlyHisGlu 109
           ||||:||||:||||:
Db   4000 GAGTCGGAAGCAGCCACCCCAGTGCACAGGTGCTCAGAAATCCTC------------ 4041

Qy   110 SerAlaGlnProTyrAlaLysHisMetLysProSerValGluProAlaProProValSer 129
                               :::|||||||||:||||||||||:|||
Db   4042 ---GCCGAGAAACTGGAACTCGACGAGACCTAGAAGACCAACCATTGCCCGGGGTGCCA 4098

Qy   130 ArgAspThrPheSerTyrMetGlyAspPheValValValTyrThrAspGlyCysCys--- 148
                                                  ||||||||||||||||||
Db   4099 ----------------------------------ACCTGTATACAGACAGCGGTAGCAGTTTC 4125

Qy   149 SerSerAsnGlyArgArgLysProArgAlaGlyIleGlyValTyrTrpGlyProGlyHis 168
           |||:||||| ||||||
Db   4126 ATCACGGAAGGTAAACGGAGA-------------------------------GCAGGGGCC 4155

Qy   169 ProLeuAsnValGlyIleArg---------------LeuProGlyArgGlnThrAsn 182
           |||:||| :::
Db   4156 CCGATCGTAGATGGCAAGCCGACGGTATGGGCTAGCAGCCTGCCAGAAGGTACGTCAGCC 4215

Qy   183 GlnArgAlaGluLeuHisAlaAlaCysLysLysAlaIleGluGlnAlaLysThrGlnAsnIle 202
           :|||::|:|::|:   ::|:|||
Db   4216 CAGAAGGCTGAACTAGTATAGCCTTGACGCAGGCATTACGCCTGGCGAAGGAAAAAACATC 4275

Qy   203 AsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsnTrpVal 222
                           :::||| || ||||||:
Db   4276 AAC-------ATCTACACGGACAGCAGTATGCTTTT-----GCCACTGCTCATATT 4320

Qy   223 GlnGly------TrpLysLysAsnGlyTrpLysThrSerAlaGlyLysLysGluValIleAsn 240
                  :::||| |||||||||||:|:|:|:|::
Db   4321 CATGGGCAATATATAAGCAGAGAGGGGGTGCTCACTTCTGCTGAAAAGATATCAAAAAC 4380

Qy   241 LysGluAspPheValAla---LeuGluLeuThrGlnLeuArgGlyMetAspIleGlnTrpMet 259
           ||||:| |||||||
Db   4369 ....

```
;            MOLECULE TYPE:  cDNA
US-08-766-528-1

Alignment Scores:
Pred. No.:            1.31e-06            Length:         8060
Score:              157.00               Matches:          52
Percent Similarity:  50.31%              Conservative:     29
Best Local Similarity: 32.30%            Mismatches:       50
Query Match:         10.16%              Indels:           30
DB:                  3                   Gaps:              9

US-10-054-313-1 (1-286) x US-08-766-528-1 (1-8060)

Qy   137 GlyAspPheValValValTyrThrAspGlyCysCysSerSer--------AsnGlyArg 153
                   ||:::| |||||| ||||||||||||||                    :|::|
Db  6585 GGAGAAGTGCTAACCTGGTTCACTCACGGA-------AGAGCTATGTGGTGGAAGTAAG 6638

Qy   154 ArgLysProArgAla----------GlyIleGlyValTyrTrpGlyProGlyHisPro 169
                   ||||                    |||
Db  6639 AGGATGTGCTGGGGCGCGGTGGTGCACCCGGCACGATCTGGGCCAGCAGC------- 6692

Qy   170 LeuAsnValGlyIleArgLeuProGlyArgGlnThrArgAsnGlnArgAlaGluIleHisAla 189
                                                        |:  ||||||||
Db  6693 ---------------------------CTGCCGGGAGGAACTTCAGCACCAAAAGGCTGAGTCATGGCC 6734

Qy   190 AlaCysLysAlaIleGluGluGlnIleAlaLysThrGlnAsnIleAsnLysAsnLysLeuTyrThr 209
                   :|:|:: ||||:|  :::|||::|||             :|:|:||||:
Db  6735 CTCACCCAAGCTTTGCGGCTGGCCGAAGGGAAATCCATAAAC-----------ATTTATACG 6785

Qy   210 AspSerMetPheThrIleAsnGlyIleThrAsnTrpValGlnGly------TrpLysLys 227
                   |||              |||:|||                    :::
Db  6786 GACAGCAGGTATGCCTTT-----GCGACTGCACACGTACATGGGGCCATCTATAAACAA 6839

Qy   228 AsnGlyTrpLysThrSerAlaGlyIleAspPheValAla--- 246
                   |||||| :|  |:|      |:
Db  6840 AGGGGTGGTTGCTTACCTCAGCAGGGAGGGAAATAAAGAGGAAATTCTAAGCCTA 6899

Qy   247 LeuGluArgLeuThrGlnIleGlyMetAspIleGlnTrpMetHisValProGlyHisSerGly 266
                   ||||| ::  |||                |:  :|
Db  6900 TTAGAGCCGTACATTTACCAAAAGGCTAGCTATTATACACTGTCCTGGACATCAGAAA 6959

Qy   267 ---------PheIleGlyAsnGlyGluAlaAspArgLeuAlaGluGlyLysAla 281
                                |||
Db  6960 GCTAAAGATCTCATATCCAGAGGAAACCAGATGGCTGACCCGGTTGCCAAGCAGGCAGCC 7019

Qy   282 Lys 282
         :::
Db  7020 CAG 7022
```

RESULT 14
US-08-766-528-1
; Sequence 1, Application US/08766528
Patent No. 6190861
GENERAL INFORMATION:
     APPLICANT: Jay A. Fishman
     TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
     TITLE OF INVENTION: AND METHODS OF USE
     NUMBER OF SEQUENCES: 74
     CURRENT APPLICATION DATA:
       APPLICATION NUMBER: US/08/766,528
       FILING DATE:
     PRIOR APPLICATION DATA:
       APPLICATION NUMBER: US 08/572,645
       FILING DATE: 14-DEC-1995
     CORRESPONDENCE ADDRESS:
       ADDRESSEE: LAHIVE & COCKFIELD, LLP
       STREET: 60 State Street
       CITY: Boston
       STATE: Massachusetts
       COUNTRY: USA
       ZIP: 02109-1875
     COMPUTER READABLE FORM:
       MEDIUM TYPE: Floppy disk
       COMPUTER: IBM PC compatible
       OPERATING SYSTEM: PC-DOS/MS-DOS
       SOFTWARE: PatentIn Release #1.0, Version #1.25
     ATTORNEY/AGENT INFORMATION:
       NAME: Louis Myers
       REGISTRATION NUMBER: 35,965
       REFERENCE/DOCKET NUMBER: MGP-038CP
     TELECOMMUNICATION INFORMATION:
       TELEPHONE: (617)227-7400
       TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 1:
     SEQUENCE CHARACTERISTICS:
       LENGTH: 8060 base pairs
       TYPE: nucleic acid
       STRANDEDNESS: single
       TOPOLOGY: linear

RESULT 15
US-08-258-420-13/c
; Sequence 13, Application US/08258420
; Patent No. 5710037
GENERAL INFORMATION:
     APPLICANT: Niehuis, Arthur W.
     APPLICANT: Vanin, Elio F.
     TITLE OF INVENTION: No. 5710037el Retroviral Envelope and LTR and Retroviral Vect
     NUMBER OF SEQUENCES: 14
     CORRESPONDENCE ADDRESS:
       ADDRESEE: Carella, Byrne, Bain, Gilfillan,Cecchi, Stewart & Olstein
       STREET: 6 Becker Farm Road
       CITY: Roseland
       STATE: New Jersey
       COUNTRY: USA
       ZIP: 07068
     COMPUTER READABLE FORM:
       MEDIUM TYPE: 3.5 inch diskette
       COMPUTER: IBM PS/2
       OPERATING SYSTEM: MS-DOS
       SOFTWARE: WordPerfect 5.1
     CURRENT APPLICATION DATA:
       APPLICATION NUMBER: US/08/258,420
       FILING DATE: 10-JUN-1994

```
Qy   149 SerSerAsnGlyArgArgLysProArgAlaGlyIleGlyValTyrTrpGlyProGlyHis 168
                   :::   |||:||:|                            --------GCAGGGGCC 4525
Db  4496 ATCACGGAAGGTAAACGGAGA-------

Qy   169 ProLeuAsnValGlyIleArg----------LeuProGlyArgGlnThrArgAsn 182
                   |||||:||                       :::
Db  4526 CCGATCGTAGATGGCAAGCGGCACGGTATGGGCTAGCAGCCTGCCAGAAGGTACGTCAGCC 4585

Qy   183 GlnArgAlaGluIleHisAlaAlaCysLysAlaIleGluGluGlnIleAlaLysThrGlnAsnIle 202
                   :::|||:| ||| :||  |:||:|
Db  4586 CAGAAGGCTGAACTAGTAGTGCCTTGACGCGCAGGCATTACCCCTGCCGGAGGAAAAAACATC 4645

Qy   203 AsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsnTrpVal 222
                   |||       :|||||      ||||||||
Db  4646 AAC---------ATCTACACGGACACAGGATGCAGGTATGCTTT-----GCCACTGCTCATATT 4690

Qy   223 GlnGly------TrpLysLysAsnGlyTrpLysThrSerAlaGlyIleAspPheValAlaIleAsn 240
                   :::             ||:::|:||              |||:||
Db  4691 CATGGGCAATATATAAGCAGAGGGGCGGCTGCTCACTTCTGCTGGAAAAGATATCAAAAC 4750

Qy   241 LysGluAspPheValAla---LeuGluArgLeuThrGlnIleGlyMetAspIleGlnTrpMet 259
                   ||| :||||||          :::
Db  4751 AAAGAGGAAATTTGGCCCTGCTAGAGCCCATCCATCTCCCTAGGCGGGTGCCATTATC 4810

Qy   260 HisProGlyHisSerGly---------PheIleGlyAsnGluGluAlaAsp 274
                   :::                        |||
Db  4811 CACTGTCCTGGCCACCAGAGGGAGTAACCCTGTGGCCACTGGGAGACCGGAGGGGCCGAC 4870

Qy   275 ArgLeuAlaArgGluGlyLysAla 281
                   :::|:::|||
Db  4871 GAGGCTGCAAAGCAAGCCGCC 4891
```

```
    CLASSIFICATION:  435
PRIOR APPLICATION DATA:
    APPLICATION NUMBER:
    FILING DATE:
ATTORNEY/AGENT INFORMATION:
    NAME:  Olstein, Elliot M.
    REGISTRATION NUMBER:  24,025
    REFERENCE/DOCKET NUMBER:  271010-208
TELECOMMUNICATION INFORMATION:
    TELEPHONE:  201-994-1700
    TELEFAX:   201-994-1744
INFORMATION FOR SEQ ID NO:  13:
SEQUENCE CHARACTERISTICS:
    LENGTH:  8202 bases
    TYPE:  nucleic acid
    STRANDEDNESS:  single
    TOPOLOGY:  linear
MOLECULE TYPE:  genomic DNA
FEATURE:
    NAME/KEY:  viral genome
US-08-258-420-13
```

```
Alignment Scores:
Pred. No.:                             Length:          8202
Score:             1.74e-06            Matches:           65
Percent Similarity:  156.00            Conservative:      30
Best Local Similarity:  41.30%         Mismatches:        68
Query Match:       28.26%              Indels:            67
DB:                10.09%              Gaps:              12
```

US-10-054-313-1 (1-286) x US-08-258-420-13 (1-8202)

```
Qy    95 ProGlyLysArgLeuArgGluProLeuAspGlyAspGly----------HisGluSer 110
            |||                                     |||
Db  3805 CCGGATGACCCATTACCAAGCCCTGCTCCTGGACACGGACCCGGGTCCAGTTCGGGCCAGT 3746

Qy   111 AlaGlnProTyrAlaLysHisMetLysProSerVal---------------------- 122
            :||    :::
Db  3745 AGTGGCCCTAAATCCAGCTACGCGCTCCCTCCCTGCCTGAGGAGGGGCTGCAACATGACTG 3686

Qy   123 --------------------------------------GluProAlaProPro 127
                                                |||||||||||
Db  3685 CCTTGACATCTTGGCTGAGCCCACGGAACTAGATCAGATCTTACGGACCAGCCCCTCCC 3626

Qy   128 ValSerArg-AspThrPheSerTyrMetGlyAspPheValValValTyrThrAspGlyCy 147
            ||||||                                     ||||||||||
Db  3625 AGACGGCCGACCACCTGG--------------------------TACACGGATGGGGAG 3593

Qy   147 sCys--SerSerAsnGlyArgArgLysProArgAlaGlyIle---------GlyVa 162
            |||                                        :::
Db  3592 CAGCTTCCTGCAAGAAGGGCAGCGTAAGGCCGAGCAGCGGTGACCACTGAGACTGAGGT 3533

Qy   162 lTyrTrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGlnThrAs 182
            ||||||                                 ||||||
Db  3532 AATCTGGGCCAGGGCA--------------------TTGCCAGCCGGGACATCGGC 3497

Qy   182 nGlnArgAlaGlyIleHisAlaAlaCysLysLysAlaIleGluGlnAlaLysThrGlnAsnIl 202
            ||||||::|||::::                    ::||::|||||||
Db  3496 CCAAAGAGCTGAACTGATAGGCTCACCCCAAGACCCTAAAGATGGCAGAAGGTAAGAAGT 3437

Qy   202 eAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsnTrpVa 222
            :|||                    :||||||||
Db  3436 AAAT---------GTTTATACTGATAGCCGTTACGCTTTT-----GCCACCGCCCATAT 3392

Qy   222 lGlnGly------TrpLysAsnGlyTrpLysThrSerAlaGlyLysGluValIleAs 240
            ||||          :::::::::||||||||
Db  3391 TCATGGAGAAATATACAGAAGGGCGCGGGTTGCTCACATCAGAAGGAAAAGAGATCAAGAA 3332

Qy   240 nLysGluAspPheValAlaLeuGluArg----------LeuThrGlnGlyMetAspIleGl 257
            ||:::|::::::   |::::|||                |||   :::||:
Db  3331 CAAGGACGAGATCTTAGCCCTACTAAGGCTCTCTTCTTGCCAAAAGACTTAGCATA-- 3274

Qy   257 nTrpMetHisValProGlyHis----------SerGlyPheIleGlyAsnGluGl 272
            |||
Db  3273 ----ATTCATTGCCCGGGACATCAAAAAGGAAACAGCGCGCAGAGGGCCAGGGGCCAACCGGAT 3218

Qy   272 uAlaAspArgLeuAlaArgGluGlyAla 281
            |||||       :::|||||
Db  3217 GGCCGACCAAGCGGCCCAGAGAAGTAGCC 3190
```

Search completed: December 17, 2003, 21:29:24
Job time : 1091 secs

```
                    GenCore version 5.1.6
           Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on:    December 17, 2003, 20:35:09 ; Search time 365 Seconds
                                    (without alignments)
                              2610.987 Million cell updates/sec

Title:          US-10-054-313-1
Perfect score:  1546
Sequence:       1 MSWLFLAHRVALAALPCRR..........FIGNEEADRLAREGAKQSED 286

Scoring table:  BLOSUM62
           Xgapop  10.0 , Xgapext  0.5
           Ygapop  10.0 , Ygapext  0.5
           Fgapop   6.0 , Fgapext  7.0
           Delop    6.0 , Delext   7.0

Searched:       2221978 seqs, 1661101734 residues

Total number of hits satisfying chosen parameters:        4423956

Minimum DB seq length:  0
Maximum DB seq length:  2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
  -MODEL=frame+_p2n.model -DEV=xlh
  -Q=/cgn2_1/USPTO_spool/US10054313/runat_17122003_150747_24333/app_query.fasta_1.455
  -DB=Published_Applications_NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
  -LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
  -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
  -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
  -MAXLEN=2000000000 -USER=US10054313 @CGN_1_1_353 @runat_17122003_150747_24333
  -NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
  -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
  -FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :     Published_Applications_NA:*
     1:  /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
     2:  /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
     3:  /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
     4:  /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
     5:  /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
     6:  /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
     7:  /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
     8:  /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
     9:  /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
    10:  /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
    11:  /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
    12:  /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
    13:  /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
    14:  /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
    15:  /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
    16:  /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
    17:  /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
    18:  /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

                         SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result                                                              *
No.    Score Match Length DB  ID                          Description
------ -----------------------------------------------------------------
```

Result No.	Score	Match	Length	DB	ID	Description
1	1526	98.7	1807	15	US-10-198-846-12889	Sequence 12889, A
2	549	35.5	310	10	US-09-796-692-9423	Sequence 9423, Ap
3	549	35.5	310	10	US-10-040-862-9423	Sequence 9423, Ap
4	541	35.0	310	15	US-09-796-692-7011	Sequence 7011, Ap
5	541	35.0	310	10	US-10-040-862-7011	Sequence 7011, Ap
6	468	30.3	436	15	US-09-918-995-29506	Sequence 29506, A
7	450	29.1	1279	13	US-09-814-353-19582	Sequence 19582, A
8	434	28.1	473	10	US-09-960-352-6980	Sequence 6980, Ap
9	399	25.8	407	10	US-09-960-352-3900	Sequence 3900, Ap
10	364	23.5	764	9	US-09-910-941-431	Sequence 431, App
11	240	15.5	499	13	US-09-814-353-13698	Sequence 13698, A
12	240	15.5	501	13	US-09-814-353-943	Sequence 943, App
13	240	15.5	501	13	US-09-814-353-7313	Sequence 7313, App
14	236.5	15.3	462	15	US-10-260-877-71	Sequence 71, Appli
15	236.5	15.3	1830121	15	US-10-329-960-1	Sequence 1, Appli
16	235.5	15.2	801	11	US-09-975-719-154	Sequence 154, App
17	235.5	15.2	2048	11	US-09-975-719-136	Sequence 136, App
18	193.5	12.5	9025608	15	US-10-156-761-11	Sequence 1, Appli
19	176.5	11.4	708	15	US-10-156-761-1127	Sequence 1127, Ap
20	154	10.0	8332	9	US-09-006-298-1	Sequence 1, Appli
21	152	9.8	8323	9	US-09-970-597-2	Sequence 2, Appli
22	150	9.7	185548	15	US-10-175-523-62	Sequence 62, Appl
23	144	9.3	8655	15	US-10-175-523-85	Sequence 85, Appl
24	141.5	9.2	640681	10	US-09-790-998-1	Sequence 1, Appli
25	140.5	9.1	13049	9	US-09-231-235-1	Sequence 1, Appli
26	140.5	9.1	13049	9	US-09-797-518A-1	Sequence 1, Appli
27	140.5	9.1	13049	11	US-09-996-073-1	Sequence 1, Appli
28	140.5	9.1	13049	11	US-09-872-696A-1	Sequence 1, Appli
29	137.5	8.9	4427	15	US-10-097-002-15	Sequence 15, Appl
30	137.5	8.9	4427	15	US-10-097-002-17	Sequence 17, Appl
31	137	8.9	2151	9	US-09-845-157-1	Sequence 1, Appli
32	136	8.8	215980	15	US-09-972-546-16	Sequence 16, Appl
33	134.5	8.7	9464	9	US-09-738-847-1	Sequence 1, Appli
34	133.5	8.6	10006	13	US-09-827-688-3	Sequence 3, Appli
35	133.5	8.6	10372	13	US-10-279-992A-1	Sequence 1, Appli
36	132.5	8.6	9876	13	US-10-279-992A-3	Sequence 3, Appli
37	132.5	8.6	9984	13	US-10-279-992A-5	Sequence 5, Appli
38	130.5	8.4	10274	13	US-09-827-688-15	Sequence 15, Appl
39	129.5	8.4	447	10	US-09-998-599-868	Sequence 868, App
40	129.5	8.4	511	10	US-09-998-599-1016	Sequence 1016, Ap
41	129.5	8.4	539	10	US-09-998-599-853	Sequence 853, App
42	129.5	8.4	540	10	US-09-998-599-820	Sequence 820, APP
43	129.5	8.4	570	10	US-09-998-599-385	Sequence 385, APP
44	129.5	8.4	571	10	US-09-998-599-1299	Sequence 1299, Ap
45	129.5	8.4	584	10	US-09-998-599-1426	Sequence 1426, Ap

```
                                          ALIGNMENTS

RESULT 1
US-10-198-846-12889
;  Sequence 12889, Application US/10198846
;  Publication No. US2003009974A1
;  GENERAL INFORMATION:
;  APPLICANT: Lillie, James
;  APPLICANT: Xu, Yongyao
;  APPLICANT: Wang, Youzhen
;  APPLICANT: Steinmann, Kathleen
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
;  TITLE OF INVENTION:  FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
;  TITLE OF INVENTION: THERAPY OF BREAST CANCER
;  FILE REFERENCE: MRI-049
;  CURRENT APPLICATION NUMBER: US/10/198,846
;  CURRENT FILING DATE: 2002-07-18
;  PRIOR APPLICATION NUMBER: 60/306,220
;  PRIOR FILING DATE: 2001-07-18
;  NUMBER OF SEQ ID NOS: 14084
;  SOFTWARE: FastSEQ for Windows Version 4.0
;  SEQ ID NO 12889
;  LENGTH: 1807
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
```

```
;  NAME/KEY: misc_feature
;  LOCATION: 1802, 1803, 1804, 1805, 1806, 1807
;  OTHER INFORMATION: n = A,T,C or G
US-10-198-846-12889

Alignment Scores:
Pred. No.:            3.41e-171        Length:           1807
Score:               1525.00           Matches:           282
Percent Similarity:  99.30%            Conservative:        2
Best Local Similarity: 98.60%          Mismatches:          2
Query Match:         98.71%            Indels:              0
DB:                                    Gaps:                0

US-10-054-313-1 (1-286) x US-10-198-846-12889 (1-1807)

Qy    1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaAlaLeuProCysArgArg 20
Db  112 ATGAGCTGGCTTCTGTTCCTGGCCCACAGAGTCGCCTTGGCCGCCTTGCCCTGCCGCCGC 171

Qy   21 GlySerArgGlyPheGlyMetPheTyrArgAlaValArgArgGlyArgLysThrGlyValPhe 40
Db  172 GGCTCTCGCGCGGTTCGGGATGTTCTATCGCGTGAGAGGGGCCCGCAAGACCGGGGTCTTT 231

Qy   41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys 60
Db  232 CTGACCTGGAATGAGTGCAGAGCACAGGTGGACCCGGTTTCCTGCTGCCAGATTTAAGAAG 291

Qy   61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80
Db  292 TTTGCCACAGAGGATGAGGCCTGGGCCTTTGTCAGGAAATCTGCAAGCCCGGAAGTTTCA 351

Qy   81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100
Db  352 GAAGGGCATGAAAATCAACATGGACAAGAATCGGACAAGAATCGGAGGGCAAAGCCAGAGAGCCACTCCGT 411

Qy  101 GluProLeuAspGlyHisAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120
Db  412 GAGCCACTGGATGGATGGAGATGGACATGGACCCGTATGCGAAAGCACATGAAGCCG 471

Qy  121 SerValGluProAlaProProValSerArgArgAspThrPheSerTyrMetGlyAspPheVal 140
Db  472 AGCGTGGAGCCCGCGCCCTCCAGTTAGCAGAGACACGTTTCCTACATGGGAGACTTCGTC 531

Qy  141 ValValTyrThrAspGlyTyrCysSerSerSerAsnGlyArgArgLysProArgAlaGlyIle 160
Db  532 GTCGTCTACACTGATGGCTGCTGCTCCAGTAATGGGCGTAGAAGGCCGCGAGCAGGAATC 591

Qy  161 GlyValTyrThrPheGlyHisProGlyAsnValGlyIleArgLeuProGlyArgGln 180
Db  592 GGCGTTTACTGGGGGCCAGGGCCCATCTTTAAATGTAGGCATTAGACTTCCTGGGGCCAG 651

Qy  181 ThrAsnGlnArgAlaAlaLeuGlyIleHisLeuAlaLeuAlaCysLysLysAlaLysThrGln 200
Db  652 ACAAACCAAAGAGCGGAAATTCATGCAGCCTGCAAAGCCATTGAACAAGCAAAGACTTCA 711

Qy  201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAspGlyIleThrAsn 220
Db  712 AACATCAATAAACTGGTTCTGTATACAGACAGTATGTTTACGATAAATGTATAACTAAC 771

Qy  221 TrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGluValIleAsn 240
Db  772 TGGGTTCAAGGTTGGAAGAAAAATGGGTGGAAGACAAGTGCAGGGAAAGAGGTGATCAAC 831

Qy  241 LysGluAspPheValAlaLeuArgLeuGlyLysLeuThrGlnGlyMetAspIleGlnTrpMetHis 260
Db  832 AAAGAGGACTTTGTGGCACTGAGAGCTTACCCAGGGATGGACATTCAGTGGATGCAT 891

Qy  261 ValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgGluGlyGly 280
Db  892 GTTCCCGGTCATTCGGGATTTATAGGCAATGAAGAAGCTGCACAGATTAGCCAGAGAGGA 951

Qy  281 AlaLysGlnSerGlnAspAsp 286
Db  952 GCTAACAAATCGGAAGAC 969
```

```
Db  952 GCTAACAAATCGGAAGAC 969

RESULT 2
US-09-796-692-9423/c
;  Sequence 9423, Application US/09796692
;  Publication No. US20020198362A1
;  GENERAL INFORMATION:
;  APPLICANT: Gaiger, Alexander
;  APPLICANT: Algate, Paul A.
;  APPLICANT: Mannion, Jane
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
;  TITLE OF INVENTION: HEMATOLOGICAL MALIGNANCIES
;  FILE REFERENCE: 2077.001200
;  CURRENT APPLICATION NUMBER: US/09/796,692
;  CURRENT FILING DATE:  2001-03-01
;  PRIOR APPLICATION NUMBER: 60/186,126
;  PRIOR FILING DATE: 2000-03-01
;  PRIOR APPLICATION NUMBER: 60/190,479
;  PRIOR FILING DATE: 2000-03-17
;  PRIOR APPLICATION NUMBER: 60/200,545
;  PRIOR FILING DATE: 2000-04-27
;  PRIOR APPLICATION NUMBER: 60/200,303
;  PRIOR FILING DATE: 2000-04-28
;  PRIOR APPLICATION NUMBER: 60/200,779
;  PRIOR FILING DATE: 2000-04-28
;  PRIOR APPLICATION NUMBER: 60/200,999
;  PRIOR FILING DATE: 2000-05-01
;  PRIOR APPLICATION NUMBER: 60/202,084
;  PRIOR FILING DATE: 2000-05-04
;  PRIOR APPLICATION NUMBER: 60/206,201
;  PRIOR FILING DATE: 2000-05-22
;  PRIOR APPLICATION NUMBER: 60/218,950
;  PRIOR FILING DATE: 2000-07-14
;  PRIOR APPLICATION NUMBER: 60/222,903
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: 60/223,416
;  PRIOR FILING DATE: 2000-08-04
;  PRIOR APPLICATION NUMBER: 60/223,378
;  PRIOR FILING DATE: 2000-08-07
;  NUMBER OF SEQ ID NOS: 9597
;  SOFTWARE: FastSEQ for Windows Version 3.0
;  SEQ ID NO 9423
;  LENGTH: 310
;  TYPE: DNA
;  ORGANISM: Homo sapiens
US-09-796-692-9423

Alignment Scores:
Pred. No.:             4.13e-56        Length:            310
Score:                549.00           Matches:            98
Percent Similarity:   98.04%           Conservative:        2
Best Local Similarity: 96.08%          Mismatches:           2
Query Match:          35.51%           Indels:              0
DB:                                    Gaps:                0

US-10-054-313-1 (1-286) x US-09-796-692-9423 (1-310)

Qy   68 TrpAlaPheValArgLysSerAlaSerProGluValSerGluGlyHisGluAsnGlnHis 87
Db  308 TGGGCCTTTGTCAGGAAATCTGCAAGCCCGGAGTTTCAGAAGGGCATGAAAATCAACAT 249

Qy   88 GlyGlnGluSerGluAlaLysProGlyLysArgLeuArgGluProLeuAspGlyAspGly 107
Db  248 GGACAAGAATCGGAGGGCAAAGCCAGAGAGCCACTCCGTGAGCCACTGGATGGATGGA 189

Qy  108 HisGluSerAlaGlnProTyrAlaLysHisMetLysProSerValGluProAlaProPro 127
Db  188 CATGAAAGCGCAGAGCCGTATGCAAAGCACATGAAGCCGAGCGTGGAGCCGGCGCCTCCA 129

Qy  128 ValSerArgArgAspThrPheSerTyrMetGlyAspPheValValTyrThrAspGlyTyrCys 147
Db  128 GTTAGCAGAGACACGTTTTCCTACATGGGAGACTTCGTCGTCGTCTACACTGATGGCTGC 69
```

```
Qy  148 CysSerSerAsnGlyArgArgLysProArgAlaGlyIleGlyValTyrTrpGlyProGly 167
            ||||||||||||||||||:::||||||||||||||||||||||||||||||||||||||
Db   68 TGCTCCAGTAATGGGCGTAGAAGGCCGCGAGCAGGAATCGGCGTTTACTGGGGGCCAGGC 9

Qy  168 HisPro 169
            ||||||
Db    8 CATCCT 3
```

RESULT 3
US-10-040-862-9423/c
; Sequence 9423, Application US/10040862
; Publication No. US20030078396A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Retter, Marc
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
; TITLE OF INVENTION: Hematological Malignancies
; FILE REFERENCE: 014058-0135200S
; CURRENT APPLICATION NUMBER: US/10/040,862
; CURRENT FILING DATE: 2001-11-06
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: US 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: US 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: US 60/223,378
; PRIOR FILING DATE: 2000-08-07
; PRIOR APPLICATION NUMBER: US 09/796,692
; PRIOR FILING DATE: 2001-03-01
; NUMBER OF SEQ ID NOS: 10467
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 9423
; LENGTH: 310
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-040-862-9423

Alignment Scores:
Pred. No.:
Score: 4.13e-56
                        549.00
Percent Similarity: 98.04%   Length: 310
Best Local Similarity: 96.08%   Matches: 98
Query Match: 35.51%   Conservative: 2
DB: 15   Mismatches: 2
                                         Indels: 0
                                         Gaps: 0

US-10-054-313-1 (1-286) x US-10-040-862-9423 (1-310)

Qy   68 TrpAlaPheValArgLysSerAlaSerProGluValSerGluProGluHisGluAsnGlnHis 87
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 308 TGGGCCTTTGTCAGGAAATCGGCAAGCCCCGGAAGTTTCAGGAAGGGCATGAAAATCAACAT 249

Qy   88 GlyGlnGluSerGluAlaLysProGlyLysArgLeuArgGluArgLeuAspGlyAspGly 107
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db  248 GGACAAGAATCCGAGGAAGCCAGCAAGCGACTCCGTGAGCCACTGGATCGAGATGGA 189

Qy  108 HisGluSerAlaGlnProTyrAlaLysHisMetLysProSerValGluProAlaProPro 127
            ||||||||:::|||||||||||||||||||||||||||||||||||||||||||||||
Db  188 CATGAAAGCGAGAGCCGTATGCAAAGCACATGAAGCCGAGCCGTGAGCCGCGCCTCCA 129

Qy  128 ValSerArgAspThrPheSerTyrMetGlyAspPheValValTyrThrAspGlyCys 147
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  128 GTTAGCAGAGACACGTTTTCCTACATGGGAGACTTCGTCGTCTACACTGATGGCTGC 69

Qy  148 CysSerSerAsnGlyArgArgLysProArgAlaGlyIleGlyValTyrTrpGlyProGly 167
            ||||||||||||||||||:::||||||||||||||||||||||||||||||||||||||
Db   68 TGCTCCAGTAATGGGCGTAGAAGGCCCCGAGCAGGAATCGGCGTTTACTGGGGGCCAGGC 9

Qy  168 HisPro 169
            ||||||
Db    8 CATCCT 3
```

RESULT 4
US-09-796-692-7011/c
; Sequence 7011, Application US/09796692
; Publication No. US2002019836 2A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAP
; TITLE OF INVENTION: HEMATOLOGICAL MALIGNANCIES
; FILE REFERENCE: 2077.001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 7011
; LENGTH: 310
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-796-692-7011

Alignment Scores:
Pred. No.:
Score: 3.69e-55
                        541.00
Percent Similarity: 97.06%   Length: 310
Best Local Similarity: 95.10%   Matches: 97
Query Match: 34.99%   Conservative: 2
DB: 10   Mismatches: 3
                                         Indels: 0
                                         Gaps: 0

US-10-054-313-1 (1-286) x US-09-796-692-7011 (1-310)
```

Alignment Scores:
Pred. No.:                                          Length:          310
Score:                       3.69e-55               Matches:          97
Percent Similarity:          97.06%                 Conservative:      2
Best Local Similarity:       95.10%                 Mismatches:        3
Query Match:                 34.99%                 Indels:            0
DB:                          541.00                 Gaps:              0

US-10-054-313-1 (1-286) x US-10-040-862-7011 (1-310)

```
Qy   68  TrpAlaPheValArgLysSerAlaSerProGluValSerGluGlyHisGluAsnGlnHis  87
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  308  TGGGCCTTGTCAGGAAATCTGCAAGCCCGGAAGTTTCAGAGAGGGCATGAAAAATCAACAT 249

Qy   88  GlyGlnGluSerGluAlaAlaLysProGlyLysArgGluArgLeuAspGlyAspGly    107
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  248  GGACAAGAATCGGAGGCGAAAGCGAGCAGCAACCAGCCAACCACTCCGTGACCACTGGAGATGGA 189

Qy  108  HisGluSerAlaGlnProTyrArgAlaLysHisMetLysProSerValGluProAlaProPro 127
         ||||||||||||||||||||||: :|||||||||||||||||||||||||||||||
Db  188  CATGAAAGCGCAGAGCCGTATGCAAAGCACATGAAAGTCGAAAGCACATGAAGTCGACGTGGAGCCGCGCCTCCA 129

Qy  128  ValSerArgAspThrPheSerTyrMetGlyAspPheValValTyrThrAspGlyCys    147
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  128  GTTAGCAGAGACACGTTTCCTACATGGAGACTTCGTCGTCGTCTACACTGATGGCTGC    69

Qy  148  CysSerSerAsnGlyTyrArgArgLysProArgAlaGlyIleGlyValTyrTrpGlyProGly 167
         |||||||||||||: :||||||||||||||||||||||||||||||||
Db   68  TGCTCCAGTAATGGGGTAGAAGGCCGTAGAAGGCCGGAGCAGGAATCGGCGTTTACTGGGGCCGGGGC 9

Qy  168  HisPro 169
         ||||||
Db    8  CATCCT 3
```

RESULT 5
US-10-040-862-7011/c
; Sequence 7011, Application US/10040862
; Publication No. US20030078396A1
GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Retter, Marc
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
; TITLE OF INVENTION: Hematological Malignancies
; FILE REFERENCE: 014058-0135
; CURRENT APPLICATION NUMBER: US/10/040,862
; CURRENT FILING DATE: 2001-11-06
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/218,950
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: US 60/222,903
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/223,416
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: US 60/223,378
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: US 09/796,692
; PRIOR FILING DATE: 2001-03-01
; NUMBER OF SEQ ID NOS: 10467
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 7011
;   LENGTH: 310
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-040-862-7011

RESULT 6
US-09-918-995-29506
; Sequence 29506, Application US/09918995
; Publication No. US20030073623A1
GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; TITLE OF INVENTION: FROM VARIOUS cDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 29506
;   LENGTH: 436
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(436)
;   OTHER INFORMATION: n = A, T, C or G
US-09-918-995-29506

Alignment Scores:
Pred. No.:   Length: 436
Score: 2.77e-46   Matches: 86
Percent Similarity: 468.00   Conservative: 3
Best Local Similarity: 100.00%   Mismatches: 0
Query Match: 96.63%   Indels: 0
DB: 30.27%   Gaps: 0
                             11

US-10-054-313-1 (1-286) x US-09-918-995-29506 (1-436)

```
Qy  198  LysThrGlnAsnIleAsnLysLeuAsnLeuValLeuTyrThrAspSerMetPheThrIleAsnGly 217
         |: :|||||||||||||||||||||||||||||||||||||||||||
Db   32  CAGACTCAAAACATCAATAAACTGGTTCTGTATCACAGACAGTATGTTTACTTTAAAATGGT 91
```

Qy   218 IleThrAsnTrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGlu 237
             ||:|:|||||||||||||||||||||||||||||||||||||||||||||||||||
Db   92 ATTTCTAACTGGGTTCAAGGTTGGAAGAAAATGGGTGGAAGACAAGTGCAGGGAAAGAG 151

Qy   238 ValIleAsnLysGlyGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGln 257
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   152 GTGATCAACAAAGGAGACTTTGTGGCACTGGAGAGGCTTACCCAGGGATGGACATTCAG 211

Qy   258 TrpMetHisValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAla 277
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   212 TGGATGCATGTTCCTGGTCATTCGGGATTTATAGGCAATGAAGAAGCTGACAGATTAGCC 271

Qy   278 ArgGlyAlaLysGlnSerGluAsp 286
             |||||||||||||||||||||||
Db   272 AGAGGAGCTAAACAATCGGAAGAC 298

RESULT 7
US-09-814-353-19582
; Sequence 19582, Application US20030165831A1
; Publication No. US20030165831A1
; GENERAL INFORMATION:
;   APPLICANT: Lee, John
;   APPLICANT: Thompson, Pamela
;   APPLICANT: Lillie, James
;   TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
;   TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
;   TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
;   FILE REFERENCE: MRI-006B
;   CURRENT APPLICATION NUMBER: US/09/814,353
;   CURRENT FILING DATE: 2001-03-21
;   PRIOR APPLICATION NUMBER: US 60/191,031
;   PRIOR FILING DATE: 2000-03-21
;   PRIOR APPLICATION NUMBER: US 60/207,124
;   PRIOR FILING DATE: 2000-05-25
;   PRIOR APPLICATION NUMBER: US 60/211,940
;   PRIOR FILING DATE: 2000-06-15
;   PRIOR APPLICATION NUMBER: US/09/216,820
;   PRIOR FILING DATE: 2000-07-07
;   PRIOR APPLICATION NUMBER: US 60/220,661
;   PRIOR FILING DATE: 2000-07-25
;   PRIOR APPLICATION NUMBER: US 60/257,672
;   PRIOR FILING DATE: 2000-12-21
;   NUMBER OF SEQ ID NOS: 22037
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 19582
;   LENGTH: 1279
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 1262, 1263, 1264, 1265, 1266, 1267, 1268, 1269, 1270, 1271,
;   LOCATION: 1272, 1273, 1274, 1275, 1276, 1277, 1278, 1279
;   OTHER INFORMATION: n = A,T,C or G
US-09-814-353-19582

Alignment Scores:
Pred. No.:   Length: 1279
Score: 1.63e-43   Matches: 110
Percent Similarity: 450.00   Conservative: 0
Best Local Similarity: 35.83%   Mismatches: 1
Query Match: 35.83%   Indels: 197
DB: 29.11%   Gaps: 1
                        13

US-10-054-313-1 (1-286) x US-09-814-353-19582 (1-1279)

Qy   171 AsnValGlyIleArgProGlyArgGlnThrAsnGlnArgAlaGluIleHisAlaAla 190
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21 AATGTAGGCATTAGACTTCCTGGGCGGCAGACAAACCAAAGAGCGGAAATTCATGCAGCC 80

Qy   191 CysLysAlaIleGluGlnGlnAlaLysThrGlnAsnIleAsnLysValLeuValTyrThrAsp 210
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   81 TGCAAAGCCATTGAACAACAAGCAAAGACTCAAAACATCAATAAAACTGGTTCTGTATACAGAC 140

(right column)

Qy   211 SerMetPheThrIleAsnGlyIle-------------------------------- 218
             |||||||||||||||||
Db   141 AGTATGTTTACGATAAATGGTAA--GCTTCACATTGATTTCTCTGTTTTTCCAGTAAC 199

Qy   218 -- 218
Db   200 TGTGAAGGGAAATGGTAGGAGGTGTTGTAACAGGCAGGACCCAAATGGGAACGGGGGG 259

Qy   218 -- 218
Db   260 ATGACATTGGTTTGTCAGGTACCGAGCAAAGAGTGAGGATTTTGGAGTCTCCCTTCTGCT 319

Qy   218 -- 218
Db   320 GCTCTGATGTTTCCACATGCTTATTTCTTTGCCAGGCACTGGAGATGCAGTCAGAAGTG 379

Qy   218 -- 218
Db   380 GAAGTGGCTCTTACTTCTAGTCTGTTGTGTGTATAAGTCACTTAAGATGGGCGTGTGACTG 439

Qy   218 -- 218
Db   440 CTTCTTTGGGAAATGCCCTGAATAGGAGCATGTAGGGGATGCTTACGAGGCTGGGGAAGG 499

Qy   218 -- 218
Db   500 CTCCTGGGAGAAATGACATCTTAAGCTGACAGCTGTGAGGTAGAATACAAGTTAGGGGAAA 559

Qy   218 -- 218
Db   560 GGGAAGAGAGAGGGTATTTCAGGCTGCTGTGCCAAAGGCCAAAGCAACTAAGGCCAGCTAAG 619

Qy   218 -- 218
Db   620 GAAGTGAAGAGTCAAGGATGGTGACTCCAGAATACAAGTGGGATCAATTATAAGAGGTAAG 679

Qy   218 -- 218
Db   680 ACTAAGGAGGTGAAAAAGGCAGGACACTCAAGATTGACTCACCAGTGATCGTGCATTTTT 739

Qy   219 ------ThrAsnTrpValGlnGlyTrpLysAsnGlyTrpLysThrSerAlaGly 235
                   ||||||||||||||||||||||||||||||||||||||||||||
Db   740 GTAGGTATAACTAACTGGGTTCAAGGTTGGAAGAAAAATGGGTGGAAGACAAGTGCAGGG 799

Qy   236 LysGlu-ValIleAsnLysGluAspPheValAlaLeuArgLeuThrGlnGlyMetAs 255
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   800 AAAGAGGGTGATCAACAAAGAGACTTTGTGGCACTGGAGAGGCTTACCCAGGGATGGA 859

Qy   255 pIleGlnTrpMetHisValProGlyHisSerGlyPheIleGlyHisGluGluAlaAspAr 275
             |||||||||||||||||||||||||||||||||||||||||||||
Db   860 CATTCAGTGGATGCATGTTCCTGGTCATTCGGGATTTATAGGCAATGAAGAAGCTGACAG 919

Qy   275 gLeuAlaArgGluGlyAla 281
             |||||||||||||||||
Db   920 ATTAGCCAGAGGAGGAGCA 938

RESULT 8
US-09-960-352-6980
; Sequence 6980, Application US/09960352
; Patent No. US2002013713919A1
; GENERAL INFORMATION:
;   APPLICANT: Warren, Wesley C.
;   APPLICANT: Tao, Nengbing
;   APPLICANT: Byatt, John C.
;   APPLICANT: Mathialagan, Nagappan
;   TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
;   TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
;   FILE REFERENCE: 16511.006/37-2(110298)C
;   CURRENT APPLICATION NUMBER: US/09/960,352
;   CURRENT FILING DATE: 2001-09-24
;   NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 6980
;   LENGTH: 473

```
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (404)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: 30-LIB3058-022-Q1-K1-H5
US-09-960-352-6980

Alignment Scores:
Pred. No.:                          Length:          473
Score:              3.4e-42          Matches:         79
Percent Similarity: 88.78%           Conservative:    8
Best Local Similarity: 80.61%        Mismatches:      11
Query Match:        28.07%           Indels:          0
DB:                                  Gaps:            0

US-10-054-313-1 (1-286) x US-09-960-352-6980 (1-473)

Qy    189 AlaAlaCysLysAlaAlaIleGluGlnAlaLysThrGlnAsnIleAsnLysLeuValLeuTyr 208
              ||||||||||||||||      |||:|||::::|||:||||||||||:|||||
Db     79 GCAGGCTGCAAAGCCATAGAGCAAGCTAAGGCTCAGGACATCACTAAGCTGGTTCTCTAC 138

Qy    209 ThrAspSerMetPheThrIleAsnGlyIleThrAsnTrpValGlnGlyTrpLysLysAsn 228
              ||||||||:|||||||||   |||||||   ||||||||||:||||||:||:|||||
Db    139 ACAGACAGATGTTTACCATCAATGGCATCACCAACCTGGTGAAGGGCTGGAAGCAGAAT 198

Qy    229 GlyTrpLysThrSerAlaGlyLysGluValIleAsnLysGluAspPheValAlaLeuGlu 248
              ||:|||    ||||||||||   |||||||||||||   ||||||||||||||||
Db    199 GGCTGGAGGACCAGCACCGGGAAGGAGGTGACCAACAAGGAGGACTTCGCGGAGCTGGAG 258

Qy    249 ArgLeuThrGlnGlyMetAspIleGlnTrpMetHisValProGlyHisSerGlyPheIle 268
              ::|||||||||||||||||||||||||||||||||||||||||||||||||||
Db    259 CGGCTGGCGCGGGACCATGACATTCAGTGGATGCACGTTCCTGGCCATTCGGGGTTTAAA 318

Qy    269 GlyAsnGluGluLeuAlaAspArgLeuAlaArgGluGlyLysGlnSerGluAsp 286
              ||||||||   |||||:||:||:|||||||    ||:||
Db    319 GGCAATGAAAGAGCGGACAGCTATCGAGAGAAGCTCGAAACACTCCCCAGAC 372

RESULT 9
US-09-960-352-3900
; Sequence 3900, Application US/09960352
; Patent No. US2002013713 9A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
SEQ ID NO 3900
; LENGTH: 407
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 17-LIB3058-058-Q1-K1-E1
US-09-960-352-3900

Alignment Scores:
Pred. No.:                          Length:          407
Score:              4.0e-38          Matches:         81
Percent Similarity: 75.81%           Conservative:    13
Best Local Similarity: 65.32%        Mismatches:      24
Query Match:        25.81%           Indels:          6
DB:                                  Gaps:            2

US-10-054-313-1 (1-286) x US-09-960-352-3900 (1-407)

Qy      1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaLeuAlaAlaLeuProCysArgArg 20
              ||||::: |||||   |||   |||||||||||||||||||||||||   ::: |||
```

```
Db     52 ATGACCCGGCTGCTGGGCGTGCACAGAGTGGCTTGGCCTTTGTGGGATGC------ 105
Qy     21 GlySerArgGlyMetPheTyrAlaValArgArgGlyArgLysThrGlyValPhe 40
              |||||||||||||||       ::::|||||||||||||||||||||||||
Db    106 ------CCCAGTCTTGCCATGTTCTATGCCGTGAGGAGGGGCCGCAAGGCCGGGGTCTTC 159
Qy     41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgArgPheProAlaAlaArgPheLysLys 60
              |||||||||:|||   ||||||||||||||||||||||||||||||||||||||||||
Db    160 CTGACCTGGAATGCAGAGCACAGTGGACCCGTTCCTGCAGCCAGATTTAAGAAG 219
Qy     61 PheAlaThrGluAspGluAlaTrpLeuThrArgLysSerAlaSerProGluValSer 80
              |||||||:|:||||||||:|||::::||  |||||||||||||||||||
Db    220 TTCGCCACAGAGGACGAGGCCTGGCCCTTTGTCAGAGAAGTCTGAAAGCCCAGTGATTCA 279
Qy     81 GluGlyHisGluAsnGlnHisGlyLeuGlyGlnGluGlnAlaLysProGlyLysArgArgLeuArg 100
              ||||  :::||||::|||||||||||||||||||||||||||||||||||  |||
Db    280 GAAGGGCAAAAACTAAACACGTAGAGAACCACAGCGAAGCGAAGCAAGCGACTCCGT 339
Qy    101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120
              ||||||||||||||||     |||:::
Db    340 GAGCCCTTGGATGAAGAT------GAAGATGCAAAGGCGTGTGCGAAGCACGTGCGGGAG 393
Qy    121 SerValGluPro 124
              ||||||
Db    394 AGCGCAGAGCCA 405

RESULT 10
US-09-910-943-431
; Sequence 431, Application US/09910943
; Patent No. US20020081610A1
; GENERAL INFORMATION:
; APPLICANT: Hemmati-Brivanlou, Ali
; APPLICANT: Altman, Curtis
; TITLE OF INVENTION: Assays and Materials for Embryonic Gene Expression
; FILE REFERENCE: 7529/1G148US1
; CURRENT APPLICATION NUMBER: US/09/910,943
; CURRENT FILING DATE: 2001-07-23
; NUMBER OF SEQ ID NOS: 742
; SOFTWARE: PatentIn version 3.1
SEQ ID NO 431
; LENGTH: 764
; TYPE: DNA
; ORGANISM: Xenopus laevis
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(764)
; OTHER INFORMATION: n may be a or g or c or t/u
US-09-910-943-431

Alignment Scores:
Pred. No.:                          Length:          764
Score:              1.36e-33         Matches:         364
Percent Similarity: 58.38%          Conservative:    83
Best Local Similarity: 47.98%        Mismatches:      18
Query Match:        23.54%          Indels:          57
DB:                                  Gaps:            15
                                     Gaps:            5

US-10-054-313-1 (1-286) x US-09-910-943-431 (1-764)

Qy     27 MetPheTyrAlaValArgArgGlyArgLysThrGlyValPheLeuThrTrpAsnGluCys 46
              ||||||||||||:||||::|||||||||||||||||:::|||
Db    249 ATGTTTTACGCTGTAAGGACTGGCCGTAAGCCTGGAGTCTACAATACGTGGGATGAATGT 308
Qy     47 ArgAlaGlnValAspArgArgPheProAlaAlaArgPheLysLysPheAlaThrGluAspGlu 66
              |||   ||||||||||||||||:::|||||||||||:|||:::
Db    309 AAAGAGCAAGTGGATCGATTCCTTTAGCAAGGTACAAGAGTTGCCTCAGAGGAGAT 368
Qy     67 AlaTrpLeuThrArgLysSerAlaSerProGluValSerGluGlyHisGluGlu 84
              |||||||| ||:||||   :||  :::||| |||   |||
Db    369 GCCTGGGAATTTGTGAGGAACACTCAGGAATCATCATCGAAAGGTTCTACTAGTGTTGAA 428
Qy     85 AsnGlnHisGlyLeuGlyGlnGluSerAlaGlnLys---------ProGly 96
              :::                                  |||
```

Db   429 ACAAAGGAGCCCCTACACAAGCTACAAAAGCTGCAGGACTGCATAATGTCATACCTCAG 488

Qy   97 LysArgLeuArgGluPro---LeuAspGlyAspGlyHisGluSerAlaGlnProTyrAla 115
           ||| :: ||| || || |||||| ||
Db   489 TCCAGAAGAAAGAAGACCACTACAGAGCTCAAGCACTGAGAAAGCATCCTCACCTAAA 548

Qy   116 LysHisMetLysProSerValGluProAlaProProValSerArgAspThrPheSerTyr 135
           ||| :::
Db   549 AGAAGCAAGCTCATTGATATCACTGATTTACCATCGTCACATAATGAACTTTTACCTAC 608

Qy   136 MetGlyAspPheValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLys 155
           |||
Db   609 ATGGGAGGACGCTCGTGTTGTATACACTGATGGCTGCTGTAGCGGAATGGGCGGGTAAAA 668

Qy   156 ProArgAlaGlyIleGlyValTyrTrp-GlyProGlyHisProLeuAsnValGlyIleAr 175
           |||||||| |||||||| ||||||||
Db   669 GCACGAGCTGGTATAGGTGTATACTGGGGCAAAGGCCTTCCTCTNAACCTTGCAGAAAA 728

Qy   175 gLeuProGlyArgGlnThrAsn------GlnArgAla 185
           |||| ||||| :::
Db   729 GCTT---GGAAGGGAGGCAAACTTAACCCAGCCGGGCT 762

RESULT 11
US-09-814-353-13698
; Sequence 13698, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
;   APPLICANT: Lee, John
;   APPLICANT: Thompson, Pamela
;   APPLICANT: Lillie, James
;   TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
;   TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
;   TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
;   FILE REFERENCE: MRI-006B
;   CURRENT APPLICATION NUMBER: US/09/814,353
;   CURRENT FILING DATE: 2001-03-21
;   PRIOR APPLICATION NUMBER: US 60/191,031
;   PRIOR FILING DATE: 2000-03-21
;   PRIOR APPLICATION NUMBER: US 60/207,124
;   PRIOR FILING DATE: 2000-05-25
;   PRIOR APPLICATION NUMBER: US 60/211,940
;   PRIOR FILING DATE: 2000-06-15
;   PRIOR APPLICATION NUMBER: US 60/216,820
;   PRIOR FILING DATE: 2000-07-07
;   PRIOR APPLICATION NUMBER: US 60/220,661
;   PRIOR FILING DATE: 2000-07-25
;   PRIOR APPLICATION NUMBER: US 60/257,672
;   PRIOR FILING DATE: 2000-12-21
;   NUMBER OF SEQ ID NOS: 22037
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 13698
;   LENGTH: 499
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-814-353-13698

Alignment Scores:
Pred. No.:
Score: 4.17e-19   Length: 499
Percent Similarity: 240.00   Matches: 47
Best Local Similarity: 100.00%   Conservative: 0
Query Match: 100.00%   Mismatches: 0
DB: 15.52%   Indels: 0
               13   Gaps: 0
US-09-814-353-13698

US-10-054-313-1 (1-286) x US-09-814-353-13698 (1-499)

Qy   171 AsnValGlyIleArgLeuProGlyArgGlnThrAsnGlnArgAlaGluIleHisAlaAla 190
           |||||||||||||||||||||||||||||||
Db   20 AATGTAGGCATTAGACTTCCTGGGCGGCAGACAAACCAAAGAGCGGAAATTCATGCAGCC 79

Qy   191 CysLysAlaIleGluGlnAlaLysThrGlnAsnIleAsnLysValLeuTyrThrAsp 210
           |||||
Db   80 TGCAAAGCCATTGAACAAGCAAAGACTCAAAACATCAATAAACTGGTTCTGTATACAGAC 139

Qy   211 SerMetPheThrIleAsnGly 217
           |||||||||||||||
Db   140 AGTATGTTTACGATAAATGGT 160

RESULT 12
US-09-814-353-943
; Sequence 943, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
;   APPLICANT: Lee, John
;   APPLICANT: Thompson, Pamela
;   APPLICANT: Lillie, James
;   TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
;   TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
;   TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
;   FILE REFERENCE: MRI-006B
;   CURRENT APPLICATION NUMBER: US/09/814,353
;   CURRENT FILING DATE: 2001-03-21
;   PRIOR APPLICATION NUMBER: US 60/191,031
;   PRIOR FILING DATE: 2000-03-21
;   PRIOR APPLICATION NUMBER: US 60/207,124
;   PRIOR FILING DATE: 2000-05-25
;   PRIOR APPLICATION NUMBER: US 60/211,940
;   PRIOR FILING DATE: 2000-06-15
;   PRIOR APPLICATION NUMBER: US 60/216,820
;   PRIOR FILING DATE: 2000-07-07
;   PRIOR APPLICATION NUMBER: US 60/220,661
;   PRIOR FILING DATE: 2000-07-25
;   PRIOR APPLICATION NUMBER: US 60/257,672
;   PRIOR FILING DATE: 2000-12-21
;   NUMBER OF SEQ ID NOS: 22037
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 943
;   LENGTH: 501
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-814-353-943

Alignment Scores:
Pred. No.:
Score: 4.19e-19   Length: 501
Percent Similarity: 240.00   Matches: 47
Best Local Similarity: 100.00%   Conservative: 0
Query Match: 100.00%   Mismatches: 0
DB: 15.52%   Indels: 0
               13   Gaps: 0
US-09-814-353-943

US-10-054-313-1 (1-286) x US-09-814-353-943 (1-501)

Qy   171 AsnValGlyIleArgLeuProGlyArgGlnThrAsnGlnArgAlaGluIleHisAlaAla 190
           |||||||||||||||||||||||||||||||
Db   19 AATGTAGGCATTAGACTTCCTGGGCGGCAGACAAACCAAAGAGCGGAAATTCATGCAGCC 78

Qy   191 CysLysAlaIleGluGlnAlaLysThrGlnAsnIleAsnLysValLeuTyrThrAsp 210
           |||||
Db   79 TGCAAAGCCATTGAACAAGCAAAGACTCAAAACATCAATAAACTGGTTCTGTATACAGAC 138

Qy   211 SerMetPheThrIleAsnGly 217
           |||||||||||||||
Db   139 AGTATGTTTACGATAAATGGT 159

RESULT 13
US-09-814-353-7313
; Sequence 7313, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
;   APPLICANT: Lee, John
;   APPLICANT: Thompson, Pamela
;   APPLICANT: Lillie, James
;   TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
;   TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
;   TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
;   FILE REFERENCE: MRI-006B

```
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7313
;   LENGTH: 501
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-814-353-7313

Alignment Scores:
Pred. No.:
Score:                    4.19e-19          Length:         501
                          240.00            Matches:         47
Percent Similarity:       100.00%           Conservative:     0
Best Local Similarity:    100.00%           Mismatches:       0
Query Match:              15.52%            Indels:           0
DB:                       13                Gaps:             0

US-10-054-313-1 (1-286) x US-09-814-353-7313 (1-501)

Qy  171 AsnValGlyIleArgLeuProGlyTyrGlnThrAsnGlnArgAlaGlyIleHisAlaAla 190
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   19 AATGTNAGGCATTAGACTTCCTGGGCGCGCAGACAAACCAAAGAGCGGAAATTCATGCAGCC 78

Qy  191 CysLysAlaIleGluGlnAlaLysThrGlnAsnIleAsnLysAsnLysValTyrThrAsp 210
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   79 TGCAAAGCCATTGAACAAGCAAAGACTCAAAACATCAATAAACTGGTTCTGTATACAGAC 138

Qy  211 SerMetPheThrIleAsnGly 217
            |||||||||||||||||
Db  139 AGTATGTTTACGATAAATGGT 159

RESULT 14
US-10-260-877-71
; Sequence 71, Application US/10260877
; Publication No. US20030021813A1
; GENERAL INFORMATION:
; APPLICANT: Abbott Laboratories
; APPLICANT: Chovan, Linda E.
; APPLICANT: Hessler, Paul E.
; APPLICANT: Reich, Karl A.
; TITLE OF INVENTION: ESSENTIAL BACTERIA GENES AND GENOME
; TITLE OF INVENTION: SCANNING IN HAEMOPHILUS INFLUENZAE FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: 'ESSNETIAL GENES'
; FILE REFERENCE: 6565 US.P1
; CURRENT APPLICATION NUMBER: US/10/260,877
; CURRENT FILING DATE: 2002-09-30
; PRIOR APPLICATION NUMBER: US/09/649,145
; PRIOR FILING DATE: 2000-08-25
; NUMBER OF SEQ ID NOS: 137
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 71
;   LENGTH: 462
;   TYPE: DNA
;   ORGANISM: H. influenzae
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(462)
;   OTHER INFORMATION: HI-0138
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (6)...(6)
;   OTHER INFORMATION: k = g or t/u at position 6
US-10-260-877-71

Alignment Scores:
Pred. No.:
Score:                    9.79e-19          Length:         462
                          236.50            Matches:         54
Percent Similarity:       49.68%            Conservative:    23
Best Local Similarity:    34.84%            Mismatches:      49
Query Match:              15.30%            Indels:          29
DB:                       15                Gaps:             4

US-10-054-313-1 (1-286) x US-10-260-877-71 (1-462)

Qy  140 ValValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgLysProArgAlaGly 159
            :: ::::::|||||||| |||  ||||
Db   13 ATTGAATTTTACTGATGGATCTTGCTTAGGTAAT-----------CCAGGGGCGGGC 60

Qy  160 IleGlyValTyrTrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArg 179
                                                     :::|||| ::
Db   61 GGAATTGGTGCCGTATTGCGTTATAAA 87

Qy  180 Gln-----------------------------------ThrAsnGlnArgAlaGluIle 187
             ||                              :|| || |||| ||
Db   88 CAACATGAAAAAACACTCTCCAAAGGCTATTCCAAACCACCAAATAAATGAAATGGAATTA 147

Qy  188 HisAlaAlaCysLysAlaIleGluGlnAlaLysThrGlnAsnIleAsnLysAsnLysValLeu 207
             ||  |||:::| | :::::::
Db  148 CGCGCTGCATTGAAGCATTAATATCATTAAAAGAACCTTGCTTG-----ATCACGCTT 201

Qy  208 TyrThrAspSerMetPheThrIleAsnGlyIleThrAsnTrpValGlnGlyTrpLysLys 227
             :::|:| |::::   :::||::::
Db  202 TATAGTGATAGCCAATATGAAAATGGCATAACCAAATGATCTTTAACTGGAAAAA 261

Qy  228 AsnGlyTrpLysTrpThrSerAlaGlyLysGluValIleAsnLysGluAspPheValAlaLeu 247
                 ::::     ::::||:::: : :: ::||::
Db  262 AATAATTGGAAAGCAAGTTCTGGAAGCCTGTAAAAAACAAGATTTATGATAGCCTTA 321

Qy  248 GluArgLeuThrGlnGlyMetAspIleGlnTrpMetHisValProGlyHisSerGlyPhe 267
             ::                    :::|||:::
Db  322 GATGAATCCATCCAACGTCATAAAATTAATTGGCATGGTAAAAGGCCATGCTGCGACAC 381

Qy  268 IleGlyAsnGluGluIleAspArgLeuAlaArgGluAlaGlyGlyAlaLys 282
              :::                       ::||:::
Db  382 AGAGAAAATGAAATTGCGATGAATTAGCAAAAAAAGGGGCAGAA 426

RESULT 15
US-10-329-960-1
; Sequence 1, Application US/10329960
; Publication No. US2003009277A1
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, Frag
; TITLE OF INVENTION: Thereof, and Uses Thereof
; FILE REFERENCE: PB186P1
; CURRENT APPLICATION NUMBER: US/10/329,960
; CURRENT FILING DATE: 2003-01-02
; PRIOR APPLICATION NUMBER: US 09/643,990
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: US 08/487,429
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/426,787
; PRIOR FILING DATE: 1995-04-21
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 1830121
;   TYPE: DNA
;   ORGANISM: Haemophilus influenzae
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (4747)..(4747)
;   OTHER INFORMATION: n equals a, t, g or c
```

FEATURE:
NAME/KEY: misc_feature
LOCATION: (9921)..(9921)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (10150)..(10150)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (29298)..(29298)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (36543)..(36543)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (36551)..(36551)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (36636)..(36636)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (40808)..(40810)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44416)..(44416)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44905)..(44905)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44975)..(44975)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (45593)..(45593)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (45732)..(45732)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (47036)..(47036)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51334)..(51334)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51602)..(51602)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51786)..(51786)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51805)..(51805)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (55369)..(55369)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:

FEATURE:
NAME/KEY: misc_feature
LOCATION: (65309)..(65309)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (65313)..(65313)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (80024)..(80024)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (100091)..(100091)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (102696)..(102696)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (105121)..(105121)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (107248)..(107248)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (117136)..(117136)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119750)..(119750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119924)..(119924)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (120038)..(120038)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (121344)..(121344)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122167)..(122167)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122336)..(122336)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131340)..(131340)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (139910)..(139910)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (140398)..(140398)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature

Search completed: December 17, 2003, 22:32:49
Job time : 953 secs

```
LOCATION: (142750)..(142750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145058)..(145058)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145171)..(145171)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145942)..(145942)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (147197)..(147197)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (150841)..(150841)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (152500)..(152500)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (152530)..(152530)
```

```
Alignment Scores:
Pred. No.:                          
Score:                       236.50
Percent Similarity:          49.68%
Best Local Similarity:       34.84%
Query Match:                 15.30%
DB:                              15

Length:                     1830121
Matches:                         54
Conservative:                    23
Mismatches:                      49
Indels:                          29
Gaps:                             4
```

```
US-10-054-313-1 (1-286) x US-10-329-960-1 (1-1830121)

Qy     140 ValValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGly 159
               :::  ::||:::||||||||||
Db  152742 ATTGAAATTTTACTGATGGATCTTGCTTAGGTAAT----------CCAGGGGCGGGC 152789

Qy     160 IleGlyValTyrTrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArg 179
                                                    :::|||    |||    :::
Db  152790 -----------------------------GGAATTGGTGCCGTATTGCGTTATAAA 152816

Qy     180 Gln--------------------------ThrAsnGlnArgAlaGluIle 187
               |||                           |||||  |||  ::::
Db  152817 CAACATGAAAAAACACTCTCCAAAGGCTATTCCAAACCACCAATAATCGAATGGAATTA 152876

Qy     188 HisAlaAlaCysLysAlaIleGluGlnAlaIleThrGlnAsnIleAsnLysLeuValLeu 207
               ::|||||||              |||:|:|
Db  152877 CGCGCTGTCATTGAAGCATTAAATACATTAAAAGAACCTTGCTG------ATCACGCTT 152930

Qy     208 TyrThrAspSerMetPheThrIleAsnGlyIleThrAsnTrpValGlnGlyTrpLysLys 227
               ||:::|:||||  :::          :::
Db  152931 TATAGTGATGATGGCCAAATATGAAAAATGGCATAACCAAATGGATCTTTAACTGGAAAAA 152990

Qy     228 AsnGlyTyrTrpLysThrSerGlyAlaGlyLysGluValIleAsnLysGluAspPheValAlaLeu 247
               |||      ||||:|||      ||:::  ::::|||::||:::
Db  152991 AATAATTGGAAAGCAAGTTCTGAAAGCCTGTAAAAAACCAAGATTTATGGATAGCCTTA 153050

Qy     248 GluArgLeuThrGlnGlyMetAspIleGlyMetHisValProGlyHisSerGlyPhe 267
               :::||                    ::||||
Db  153051 GATGAATCCATCCAACGTCATAAAATTAATTGGCAATGGGTAAAAGGCCATGCTGGACAC 153110

Qy     268 IleGlyAsnGluGluAlaAspArgLeuAlaArgGluGlyAlaGlyLys 282
               :::|||                                      ::||
Db  153111 AGAGAAAATGAAATTTCGATGAATGAATTAGCAAAAAAAGGGGCCAGAA 153155
```

Pred. No.:   1830121

```
                    Copyright    GenCore version 5.1.6
                    Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on:      December 17, 2003, 19:23:44 ; Search time 285 Seconds
                                           (without alignments)
                               2002.513 Million cell updates/sec

Title:            US-10-054-313-1
Perfect score:    1546
Sequence:         1 MSWLLFLAHRVALAALPCRR........FIGNEEADRLAREGAKQSED 286

Scoring table:    BLOSUM62
                  Xgapop 10.0 , Xgapext 0.5
                  Ygapop 10.0 , Ygapext 0.5
                  Fgapop  6.0 , Fgapext 7.0
                  Delop   6.0 , Delext  7.0

Searched:         3357611 seqs, 997755592 residues

Total number of hits satisfying chosen parameters:       6715222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :    Pending_Patents NA_New:*
    1: /cgn2_6/ptodata/1/pna/PCT_NEW_COMB.seq:*
    2: /cgn2_6/ptodata/1/pna/US06_NEW_COMB.seq:*
    3: /cgn2_6/ptodata/1/pna/US07_NEW_COMB.seq:*
    4: /cgn2_6/ptodata/1/pna/US08_NEW_COMB.seq:*
    5: /cgn2_6/ptodata/1/pna/US09_NEW_COMB.seq:*
    6: /cgn2_6/ptodata/1/pna/US10_NEW_COMB.seq:*
    7: /cgn2_6/ptodata/1/pna/US60_NEW_COMB.seq:*

Command line parameters:
-MODEL=frame+_p2n.model -DRV=xlh
-Q=/cgn2_1/USPTO_spool/US10054313/runat_17122003_150746_24307/app_query.fasta_1.455
-DB=Pending_Patents_NA_New -QFMT=fastap -SUFFIX=rnpn -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7
-USER=US10054313 @CGN_1_282 @runat 17122003_150746_24307 -NCPU=6 -ICPU=3
```

Pred. No. is the number of results predicted by chance to have to be printed,
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

C	Result No.	Score	% Query Match	Match Length	DB	ID	Description
c	1	1526	98.7	1168	7	US-60-490-890-1779	Sequence 1779, Ap
c	2	549	35.5	310	5	US-09-796-692A-9423	Sequence 9423, Ap
c	3	541	35.5	310	5	US-09-796-692A-7011	Sequence 7011, Ap
	4	419	27.1	629	7	US-60-507-481-9585	Sequence 9585, Ap
c	5	272	17.6	119211	6	US-10-649-136-40	Sequence 40, Appl
c	6	272	17.6	119211	6	US-10-672-787-40	Sequence 40, Appl
	7	272	17.6	36471	5	US-09-806-866A-1	Sequence 1, Appli
	8	229.5	14.8	2242716	5	US-09-806-866A-1068	Sequence 1068, Ap
	9	229.5	14.8	1733	5	US-09-897-516A-658	Sequence 658, App
c	10	212.5	13.7	1733	5	US-09-897-516A-659	Sequence 659, App
c	11	212.5	13.7	1733	5	US-09-897-516A-660	Sequence 660, App
	12	208.5	13.5	1592	6	US-10-684-141-41	Sequence 41, Appl
c	13	194.5	12.6	2731748	6	US-10-297-465B-1	Sequence 1, Appli
	14	171.5	11.1	1545	6	US-10-425-114A-34337	Sequence 34337, A
	15	162.5	10.5	644	6	US-10-653-047-5466	Sequence 5466, Ap
	16	156	10.1	8889	6	US-10-677-558-4	Sequence 4, Appli
	17	156	10.1	11364	6	US-10-677-558-11	Sequence 11, Appl
	18	156	10.1	11394	6	US-10-677-558-5	Sequence 5, Appli
c	19	139.5	9.0	270	5	US-09-674-546A-183	Sequence 183, App
c	20	138	8.9	270	5	US-09-674-546A-181	Sequence 181, App
c	21	138	8.9	270	5	US-09-674-546A-185	Sequence 185, App
	22	137	8.9	2151	1	PCT-US03-28802-1	Sequence 1, Appli
	23	137	8.9	2151	6	US-10-661-819-1	Sequence 1, Appli
	24	134.5	8.7	9464	6	US-10-673-023-1	Sequence 1, Appli
	25	134.5	8.7	9464	6	US-10-673-008-1	Sequence 1, Appli
	26	134.5	8.7	9464	6	US-10-702-75-1	Sequence 1, Appli
	27	134.5	8.7	9464	6	US-10-672-489-1	Sequence 1, Appli
c	28	125	8.1	8059021	5	US-09-947-914-53	Sequence 53, Appl
	29	124	8.0	8003	5	US-10-276-869A-38	Sequence 38, Appl
c	30	121	7.8	111190	7	US-60-500-315-11870	Sequence 11870, A
	31	121	7.8	184000	7	US-60-495-114-16699	Sequence 16699, A
c	32	121	7.8	1748349	5	US-09-947-914-48	Sequence 48, Appl
	33	116	7.5	40260	7	US-60-500-315-11639	Sequence 11639, A
	34	115.5	7.5	16360	1	PCT-US03-20325-1	Sequence 1, Appli
	35	115.5	7.5	17207	1	PCT-US03-20325-8	Sequence 8, Appli
	36	113.5	7.3	469219	7	US-60-495-114-16704	Sequence 16704, A
	37	113	7.3	21744	7	US-60-500-315-11634	Sequence 11634, A
	38	112.5	7.3	26021	7	US-60-495-114-16436	Sequence 16436, A
c	39	112	7.2	141158	7	US-10-719-993-6837	Sequence 6837, Ap
c	40	111.5	7.2	244289	7	US-60-495-114-16866	Sequence 16866, A
	41	111	7.2	1203	1	PCT-US03-33947-307	Sequence 307, App
	42	111	7.2	1203	1	PCT-US03-33947-1875	Sequence 1875, Ap
	43	111	7.2	25821	7	US-60-487-610-20089	Sequence 20089, A
	44	111	7.2	25821	7	US-60-485-450-12348	Sequence 12348, A
c	45	109	7.1	13831263	5	US-09-947-914-41	Sequence 41, Appl

ALIGNMENTS

```
RESULT 1
US-60-490-890-1779

; Sequence 1779, Application US/60490890
; GENERAL INFORMATION:
;  APPLICANT: Li, Martha
;  APPLICANT: Ruprow, Brent A.
;  APPLICANT: Webster, Kevin R.
;  APPLICANT: Jackson, Donald
;  APPLICANT: Wong, Tai W.
;  TITLE OF INVENTION: BIOMARKERS OF CYCLIN-DEPENDENT KINASE MODULATION
;  FILE REFERENCE: D0310 PSP
;  CURRENT APPLICATION NUMBER: US-60/490,890
;  CURRENT FILING DATE: 2003-07-29
;  NUMBER OF SEQ ID NOS: 2779
;  SOFTWARE: PatentIn version 3.2
;  SEQ ID NO 1779
;  LENGTH: 1168
;  TYPE: DNA
;  ORGANISM: Homo sapiens
US-60-490-890-1779

  Alignment Scores:
  Pred. No.:                 3.66e-172          Length:         1168
  Score:                     1526.00            Matches:        282
  Percent Similarity:        99.30%             Conservative:   2
  Best Local Similarity:     98.60%             Mismatches:     2
  Query Match:               98.71%             Indels:         0
  DB:                        7                  Gaps:           0

  US-10-054-313-1 (1-286) x US-60-490-890-1779 (1-1168)

Qy    1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaAlaLeuProCysArgArg 20

Db   82 ATGAGCTGGCTTCTGTTCCTGGCCCACAGAGTCGCCTTGGCCGCTCTGCCGTGCCGCCGC 141
```

Right-margin sequence descriptions

```
Qy    21  GlySerArgGlyPheGlyMetPheTyrAlaValArgArgGlyValArgLysThrGlyValPhe   40
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   142  GGCTCTCGCGGGTTCGGGATGTTCTATGCCGTGAGGAGGGGCGTCAAGACCGGGGTCTTT  201

Qy    41  LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys   60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   202  CTGACCTGGAATGAGTGCAGAGCACAGTGGACCGGTTCCTGCTGCCAGATTTAAGAAG  261

Qy    61  PheAlaThrGluAspGluAlaTrpAlaArgValArgSerAlaSerProGluValSer   80
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   262  TTTGCCACAGAGGATGAGGCCTGGGCCTTTGTCAGGAGAAATCTGCAAGCCCGGAAGTTTCA  321

Qy    81  GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg  100
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   322  GAAGGGCATGAAAATCAACATGGACAAGGAAAGCGAGGGCAAAGCCAGCAAGCGACTCCGT  381

Qy   101  GluProLeuAspGlyHisAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro  120
              |||||||||||||||||||   |||||:::||||||||||||||||||||||||
Db   382  GAGCCACTGGATGGACATGGACAAGCGCCAGAGCCGTATGCAAAGCACATGAAGCCG  441

Qy   121  SerValGluProAlaProProValSerProAspSerArgAspMetGlyAspPheVal  140
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   442  AGCGTGGAGCCGGCCGCCTCCAGTTAGCAGAGACACGTTTCCTACATGGGAGACTTCGTC  501

Qy   141  ValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGlyIle  160
              |||||||||||||||   ||||||:::||||||||||||||||||||||||
Db   502  GTCGTCTACACTGATGGCTGCTGCTCCCAGTAATGGCCGTAGAAGCCCGGCAGCAGGAATC  561

Qy   161  GlyValTyrTrpGlyTrpProGlyHisProGlyLeuAsnValGlyTyrIleArgGlyLyrProGlyValArgGln  180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   562  GGCGTTTACTCGGGGGCCAAGGCCATCCTTTAAATGTAGGCATTAGACTTCCTGGGCGCAG  621

Qy   181  ThrAsnGlnArgAlaGlyLeuHisAlaAlaCysLysAlaAlaIleGluGlnAsnAlaLysThrGln  200
              |||||||||||||||||||||||||||||||||||||||:::||||||||||||
Db   622  ACAAACCAAAGAGCGGGAAATTCATGCAGCCTGCAAAGCCATTGAACAACAGCAAAGACTCAA  681

Qy   201  AsnIleAsnLysAsnValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn  220
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   682  AACATCAATAAACTGGTTCGTATACAGACAGTATGTTTACGATAAAATGTATAAACTAAC  741

Qy   221  TrpValGlnGlyTyrTrpLysLysAsnGlyTyrTrpLysTrpSerArgAlaGlyLylTrpLysGluValIleAsn  240
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   742  TGGGTTCAAGGTTGGAAGAAAAAATGGGTGGAAGACAAGGTGCAGGGAAAGAGGTGATCAAC  801

Qy   241  LysGluAspPheValAlaLeuGluArgLeuThrGlnMetAspIleGlnTrpMetHis  260
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   802  AAAGAGGACTTTGTGGCACTGGAGAGGCTTACCCAGGGGATGGACATTCAGTGATGCAT  861

Qy   261  ValProGlyHisSerGlyPheIleGlyAsnGluGlnAlaArgArgLeuAlaAlaArgGluGly  280
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   862  GTTCCTGGTCATTCGGGATTTATAGGCAATGAAGAAGCTGACAGATTAGCCAGAGAAGGA  921

Qy   281  AlaLysGlnSerGluAsp  286
              ||||||||||||||||||
Db   922  GCTAAACAATCGGAAGAC  939
```

RESULT 2
US-09-796-692A-9423/c
; Sequence 9423, Application US/09796692A
; GENERAL INFORMATION:
;　APPLICANT:　Gaiger, Alexander
;　APPLICANT:　Algate, Paul A.
;　APPLICANT:　Mannion, Jane
;　APPLICANT:　Corixa Corporation
;　TITLE OF INVENTION:　Compositions and Methods for the Detection, Diagnosis and Therapy
;　TITLE OF INVENTION:　Hematological Malignancies
;　FILE REFERENCE:　014058-0135I2US
;　CURRENT APPLICATION NUMBER:　US/09/796,692A
;　CURRENT FILING DATE:　2001-03-01
;　PRIOR APPLICATION NUMBER:　US 60/186,126
;　PRIOR FILING DATE:　2000-03-01
;　PRIOR APPLICATION NUMBER:　US 60/190,479
;　PRIOR FILING DATE:　2000-03-17

;　PRIOR APPLICATION NUMBER:　US 60/200,545
;　PRIOR FILING DATE:　2000-04-27
;　PRIOR APPLICATION NUMBER:　US 60/200,303
;　PRIOR FILING DATE:　2000-04-28
;　PRIOR APPLICATION NUMBER:　US 60/200,779
;　PRIOR FILING DATE:　2000-04-28
;　PRIOR APPLICATION NUMBER:　US 60/200,999
;　PRIOR FILING DATE:　2000-05-01
;　PRIOR APPLICATION NUMBER:　US 60/202,084
;　PRIOR FILING DATE:　2000-05-04
;　PRIOR APPLICATION NUMBER:　US 60/206,201
;　PRIOR FILING DATE:　2000-05-22
;　PRIOR APPLICATION NUMBER:　US 60/218,950
;　PRIOR FILING DATE:　2000-07-14
;　PRIOR APPLICATION NUMBER:　US 60/222,903
;　PRIOR FILING DATE:　2000-08-03
;　Remaining Prior Application data removed - See File Wrapper or PALM.
;　NUMBER OF SEQ ID NOS:　9597
;　SOFTWARE:　FastSEQ for Windows Version 3.0
; SEQ ID NO 9423
;　LENGTH:　310
;　TYPE:　DNA
;　ORGANISM:　Homo sapiens
US-09-796-692A-9423

Alignment Scores:
Pred. No.:
Score:　　　　　　　　　　　　4.78e-56　　　Length:　　　　　　310
Matches:　　　　　549
Percent Similarity:　　　98.04%　　　　　Conservative:　　　98
Best Local Similarity:　96.08%　　　　　Mismatches:　　　　　2
Query Match:　　　　　　　35.51%　　　　　Indels:　　　　　　　0
DB:　　　　　　　　　　　　5　　　　　　　　Gaps:　　　　　　　　0

US-10-054-313-1 (1-286) x US-09-796-692A-9423 (1-310)

```
Qy    68  TrpAlaPheValArgLysSerAlaSerProGluValSerGluGlyHisGluAsnGlnHis   87
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   308  TGGGCCCTTGTCAGGAAATCTGCAAGCCCGGAAGTTTCAGAGGGCATGAAAATCAACAT  249

Qy    88  GlyGlnGluSerGluAlaLysProGlyLysArgLeuArgGluProLeuAspGlyAspGly  107
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   248  GGACAAGGAAATCGGAGGCGAAAGCCAGCAAGCGACTCCGTGAGCCACTGGATGGAGATGGA  189

Qy   108  HisGluSerAlaGlnProTyrAlaLysHisMetLysProSerValGluProAlaProPro  127
              |||||||||||||||||||:::|||||||||||||||||||||||||||||||
Db   188  CATGAAAGCGCAGAGCCGTATGCAAAGCACATGAAGCCATGAGTGGAGCCGGCCGCCTCCA  129

Qy   128  ValSerArgAspThrPheSerTyrMetGlyAspPheValValValTyrThrAspGlyCys  147
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   128  GTTAGCAGAGACACGTTTCCTACATGGGAGACTTCGTCGTCGTCTACACTGATGGCTGC  69

Qy   148  CysSerSerAsnGlyArgArgLysProArgAlaGlyIleGlyValTyrTrpGlyTrpProGly  167
              ||||||||:::||||||||||||||||||||||||||||||||:::||||||||||
Db    68  TGCTCCAGTAATGGCCGTAGAAGCCCGAGCAGCAGGAATCGGCGTTTACTGGGGGCCAGGC  9

Qy   168  HisPro  169
              ||||||
Db     8  CATCCT  3
```

RESULT 3
US-09-796-692A-7011/c
; Sequence 7011, Application US/09796692A
; GENERAL INFORMATION:
;　APPLICANT:　Gaiger, Alexander
;　APPLICANT:　Algate, Paul A.
;　APPLICANT:　Mannion, Jane
;　APPLICANT:　Corixa Corporation
;　TITLE OF INVENTION:　Compositions and Methods for the Detection, Diagnosis and Therapy
;　TITLE OF INVENTION:　Hematological Malignancies
;　FILE REFERENCE:　014058-0135I2US
;　CURRENT APPLICATION NUMBER:　US/09/796,692A
;　CURRENT FILING DATE:　2001-03-01
;　PRIOR APPLICATION NUMBER:　US 60/186,126

```
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: US 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/222,903
; PRIOR FILING DATE: 2000-08-03
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSEQ for Windows Version 3.0
;
; SEQ ID NO 7011
; LENGTH: 310
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-796-692A-7011

Alignment Scores:
Pred. No.:                   4.34e-55
Score:                       541.00
Percent Similarity:          97.06%
Best Local Similarity:       95.10%
Query Match:                 34.99%
DB:                          5

US-10-054-313-1 (1-286) x US-09-796-692A-7011 (1-310)

Qy     68 TrpAlaPheValArgLysSerAlaSerProGluValSerGluGlyHisGluAsnGlnHis 87
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    308 TGGGCCCTTGTCAGGAAATCTGCAAGCCCGGAAGTTTCAGAGAGGGCATGAAAATCAACAT 249

Qy     88 GlyGlnGluSerGluAlaLysProGlyLysArgLeuArgGluProLeuAspGlyAspGly 107
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    248 GGACAAGAATCGGAGGCGAGGCGAAAGCCAGCCAGCCACTGGATGGAGGATGGA 189

Qy    108 HisGluSerAlaGlnProTyrAlaLysHisMetLysProSerValGluProAlaProPro 127
          ||||||||||||||::::|||||||||||||||||||||||||||||||||||||||||||
Db    188 CATGAAAGCGCAGAGCCGTATGCAAAGCACATGAAGTCGAGTCGAGCGTGGAGCCGGCCTCCA 129

Qy    128 ValSerArgAspThrPheSerTyrMetGlyAspPheValValValTyrThrAspGlyGlyCys 147
          ||||||||||||||||||::::|||||||||||||||||||||||||||||||||||||||
Db    128 GTTAGCAGAGACACGTTTCCTCACATGGGAGACTTCGTCGTCGTCACACTGATGGCTGC 69

Qy    148 CysSerSerAsnGlyArgArgProArgAlaGlyIleGlyValTyrTrpGlyProGly 167
          ||||||||||||::::|||::|||||||||||||||||||||||||||||||||||||
Db     68 TGCTCCAGTAATGGGCGTAGAGGCCGCGACGAGCAGAATCGGCGTTTACTGGGGGCCGGGC 9

Qy    168 HisPro 169
          ||||||
Db      8 CATCCT 3
```

```
RESULT 4
US-60-507-481-9585
; Sequence 9585, Application US/60507481
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William M
; TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION IN ANIMAL
; TITLE OF INVENTION: MODLES OF INFLAMMATORY DISEASES
; FILE REFERENCE: AM101084
; CURRENT APPLICATION NUMBER: US/60/507,481
; CURRENT FILING DATE: 2003-10-02
```

```
; NUMBER OF SEQ ID NOS: 210107
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 9585
; LENGTH: 629
; TYPE: DNA
; ORGANISM: Canis familiaris
US-60-507-481-9585

Alignment Scores:
Pred. No.:                   5.92e-40
Score:                       419.00
Percent Similarity:          84.07%
Best Local Similarity:       78.76%
Query Match:                 27.10%
DB:                          7

US-10-054-313-1 (1-286) x US-60-507-481-9585 (1-629)

Qy    178 GlyArgGlnThrAsnGln-ArgAlaGluIleHisAlaAlaCysAlaLysAlaAlaIleGluGlnAlAl 197
          |||||||||||||||||   :::|:|||||||
Db     12 GGCTGCCAGACCGACCACAGAGCAGAAATTCATGCAGCCTGCAAAGCCATCGAGCAGCA 71

Qy    197 aLysThrGlnAsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGl 217
          :::      |||  ||| ||||||||||||||||||||||
Db     72 AAGGCGGCCATACACTTTAGAAGTTAGTTCCTGTATACGACACGTATGTTACTATAAATGG 131

Qy    217 YIleThrAsnTrpValGlnGlyTrpLysLysAsn-GlyTrpLysThrSerAlaGlyLysG 237
          |||||||||||||||||||||||||||||||||||   |||||||||||||||||||||||
Db    132 TATCACTAATTGGGTTCAAGGTTCGAAGAAGAATTGGGTGGAAAACAAGCACAGGGAAAG 191

Qy    237 luValIleAsnLysGluAspPheVal-AlaLeuGlnArgLeuThrGlnGlyMetAspIle 256
          ||||||||||||||||||||||||||   |||||||||||||||||||||||||||
Db    192 AGGTGATCAACAAGGAGGACTTTGGTCCTGGAGCAGCTCACCCAGGGCATGGACATC 251

Qy    257 GlnTrpMetHisValProGlyPheIleGlyAsnGluAsnAspArgLeu 276
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db    252 CAGTGGATGCATGTTCCTGGTCATTCGGGATTTACAGGCAATGAGAGAAGCTGACAGATTA 311

Qy    277 Ala-ArgGluGlyAlaLysGlnSerGluAsp 286
          |||   |||||||||||||||||||
Db    312 GCACAGAGAGGAGCTAAGCAATCTGAAGAC 342
```

```
RESULT 5
US-10-649-136-40/c
; Sequence 40, Application US/10649136
; GENERAL INFORMATION:
; APPLICANT: LAGACE, Robert, E.
; APPLICANT: PATTERSON, Chandra
; APPLICANT: BERG, Kim, L.
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME
; FILE REFERENCE: ELITRA.025C1
; CURRENT APPLICATION NUMBER: US/10/649,136
; CURRENT FILING DATE: 2003-08-26
; PRIOR APPLICATION NUMBER: 09/596,002
; PRIOR FILING DATE: 2000-06-16
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PERL Program
; SEQ ID NO 40
; LENGTH: 119211
; TYPE: DNA
; ORGANISM: Moraxella catarrhalis
US-10-649-136-40

Alignment Scores:
Pred. No.:                   1.59e-18
Score:                       272.00
Percent Similarity:          52.23%
Best Local Similarity:       39.49%
Query Match:                 17.59%
DB:                          6

US-10-054-313-1 (1-286) x US-10-649-136-40 (1-119211)
```

Length information (right side):
```
Length:        629
Matches:       89
Conservative:  6
Mismatches:    14
Indels:        4
Gaps:          0
```

```
Length:        310
Matches:       97
Conservative:  2
Mismatches:    3
Indels:        0
Gaps:          0
```

```
Length:        119211
Matches:       62
Conservative:  20
Mismatches:    51
Indels:        24
Gaps:          5
```

Qy   136 MetGlyAspPheValValValTyrThrArgGlyCysSerSerAsnGlyArgArgLys 155
           ::: ||
Db 45182 ATGAGCCAACTTAATATCGCTATACCGACGCGGCCGTGTAAAGGCAATGGAAAACAAGGC 45123

Qy   156 ProArgAla---GlyIleGlyValTyr--- 163
           ||| |||
Db 45122 GTATCTGCGAGGCCGGTTGGGCGTGTATTTGCATTATTTCAATGCGGCATCTG 45063

Qy   164 TrpGlyProGlyHisProLeuAsnValGlyIleArgProGlyArgGlnThrAsnGln 183
           ||||
Db 45062 TGGGGC--GGTGAGCCT 45033

Db 45062 TGGGGC--GGTGAGCCT----------------GATACGACCAATAAT 45033
Qy   184 ArgAlaGluIleHisAlaAlaCysLysAlaIleGluGlnAlaLysThrGlnAsnIleAsn 203
           :::   ----ATT 44979
Db 45032 CGCATGGAATTGATGGCTGCCATCACAGCTTTGGAGGCAACGCCTGCACAG------ATT 44979

Qy   204 LysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsnTrpValGln 223
           ::: ||||| ::::
Db 44978 CCTCTGCCAACTTTGGACAGATTCAGGCTATGTTAAAGATGGCATAACTCAGTGGATTGGC 44919

Qy   224 GlyTrpLysLysAsnGlyIleTrpLysThrSerAlaGlyLysLysGluValIleLeuAsnLysGluAsp 243
           :::
Db 44918 GGTTGGAAGTTGCGTGGTTGGAAAAAAGCAGATGGCAAGCCTGTCTTAAATCAAGACCTA 44859

Qy   244 PheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHisValProGly 263
           :::::::||
Db 44858 TGGCAACGATTGGATCAGCTGACCCAAAATCGCATCATTGATTGGCAATGGATCAAAGGC 44799

Qy   264 HisSerGlyPheIleGlyValAsnGluAlaAspArgLeuAlaArgGluGly 280
           ::: || || ::::
Db 44798 CATGCTGGTCATGCTGGCAATGAGATGGGCAATGAGACCAACTGGCCAATAAAGGC 44748

RESULT 7
US-09-806-866A-1
; Sequence 1, Application US/09806866A
; GENERAL INFORMATION:
;   APPLICANT: Frazer, Claire M.
;   APPLICANT: Hickey, Erin
;   APPLICANT: Peterson, Jeremy
;   APPLICANT: Tettelin, Herve
;   APPLICANT: Venter, J. Craig
;   APPLICANT: Masignani, Vega
;   APPLICANT: Galeotti, Cesira
;   APPLICANT: Mora, Manrosa
;   APPLICANT: Ratti, Giulio
;   APPLICANT: Scarselli, Maria
;   APPLICANT: Scarlato, Vincenzo
;   APPLICANT: Rappuoli, Rino
;   APPLICANT: Pizza, Mariagratia
;   APPLICANT: Grandi, Guido
;   TITLE OF INVENTION: Neisseria Genomic Sequences And Methods Of Their Use
;   FILE REFERENCE: CHIR0313
;   CURRENT APPLICATION NUMBER: US/09/806,866A
;   CURRENT FILING DATE: 2001-09-06
;   PRIOR APPLICATION NUMBER: USSN 60/103,794
;   PRIOR FILING DATE: 1998-10-09
;   PRIOR APPLICATION NUMBER: USSN 60/132,068
;   PRIOR FILING DATE: 1999-04-30
;   PRIOR APPLICATION NUMBER: PCT/US99/25373
;   PRIOR FILING DATE: 1999-10-08
;   NUMBER OF SEQ ID NOS: 1068
;   SOFTWARE: Patentin version 3.2
; SEQ ID NO 1
;   LENGTH: 36471
;   TYPE: DNA
;   ORGANISM: Neisseria meningitidis
US-09-806-866A-1

Alignment Scores:
Pred. No.:
Score: 2.72e-14
                            229.50
Percent Similarity: 48.73%   Length: 36471
Best Local Similarity: 34.18%   Matches: 54
Query Match: 14.84%   Conservative: 23
DB: Mismatches: 56
                                          Indels: 25
                                          Gaps: 5

US-10-054-313-1 (1-286) x US-09-806-866A-1 (1-36471)

Qy   136 MetGlyAspPheValValValTyrThrAspGlyCysCysSerSerAsnGlyCysCysSerSerAsnGlyArgArgLys 155
           ::: ||
Db 30559 ATGAACCAAACCGTTTACCTTTACCTTTACCGACGCGGCCGTGCAAAGGCAAT----------- 30606
Qy   156 ProArgAlaGlyIleGlyValValTyrTrpGlyProGlyHisProGlyHisProLeuAsnValGlyIleIleArg 175

Qy   136 MetGlyAspPheValValValTyrThrArgGlyCysSerSerAsnGlyArgArgLys 155
           ::: ||
Db 45182 ATGAGCCAACTTAATATCGCTATACCGACGCGGCCGTGTAAAGGCAATGGAAAACAAGGC 45123

Qy   156 ProArgAla---GlyIleGlyValTyr--- 163
           ||| :::
Db 45122 GTATCTGCGAGGCCGGTTGGGCGTGTATTTGCATTATTTCAATGCGGCATCTG 45063

Qy   164 TrpGlyProGlyHisProLeuAsnValGlyIleArgProGlyArgGlnThrAsnGln 183
           |||| ||
Db 45062 TGGGGC--GGTGAGCCT 45033

RESULT 6
US-10-672-787-40/c
; Sequence 40, Application US/10672787
; GENERAL INFORMATION:
;   APPLICANT: LACAGE, Robert, E.
;   APPLICANT: PATTERSON, Chandra
;   APPLICANT: BERG, Kim, L.
;   TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME
;   FILE REFERENCE: ELITRA.025C1
;   CURRENT APPLICATION NUMBER: US/10/672,787
;   CURRENT FILING DATE: 2003-09-26
;   PRIOR APPLICATION NUMBER: 09/596,002
;   PRIOR FILING DATE: 2000-06-16
;   NUMBER OF SEQ ID NOS: 41
;   SOFTWARE: PERL Program
; SEQ ID NO 40
;   LENGTH: 119211
;   TYPE: DNA
;   ORGANISM: Moraxella catarrhalis
US-10-672-787-40

Alignment Scores:
Pred. No.:
Score: 1.59e-18
                            272.00
Percent Similarity: 52.23%   Length: 119211
Best Local Similarity: 39.49%   Matches: 62
Query Match: 17.59%   Conservative: 20
DB: Mismatches: 51
                                          Indels: 24
                                          Gaps: 5

US-10-054-313-1 (1-286) x US-10-672-787-40 (1-119211)

Qy   136 MetGlyAspPheValValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLys 155
           :::
Db 45182 ATGAGCCAACTTAATATCGCTATACCGACGCGGCCGTGTAAAGGCAATGGAAAACAAGGC 45123

Qy   156 ProArgAla---GlyIleGlyValTyr--- 163
           :::
Db 45122 GTATCTGCGAGGCCGGTTGGGCGTGTATTTGCATTATTTCAATGCGGCATCTG 45063

Qy   164 TrpGlyProGlyHisProLeuAsnValGlyIleArgProGlyArgGlnThrAsnGln 183
           ||||

Qy   136 MetGlyAspPheValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLys 155
         ||| ::|||||||||||||||||||||||||||||||||
Db 1651985 ATGAACCAAACCGTTTACCTTTACCTTTACACCGACGGCGCGTGCAAAGGCAAT------- 1652032

Qy   156 ProArgAlaGlyIleGlyIleGlyValTyrValTyrProGlyHisProLeuAsnValGlyIleArg 175
         ||| :||:||:|||
Db 1652033 CCCGGCGCGGGC------------GGCTGGGGC--------------GTGTTAATGCGC 1652065

Qy   176 LeuProGlyArgGln------------------------------ThrAsnGlnArgAla 185
         ||| :::
Db 1652066 TACGGTAGCCACGAAAAAGAACTTTCGGCGGCGAAGCGCAAACCACCAACCAACCGCATG 1652125

Qy   186 GluIleHisAlaAlaCysLysLysAlaIleGluGlnIleAlaIleLysThrGlnAsnIleAsnLysLeu 205
         ||| :: :: :::::: ||||
Db 1652126 GAACTGACTGCCGTCATCGAAGGACTGAAATCGCTCAAACGCCGC------TGCACCGTC 1652179

Qy   206 ValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsnTrpValGlnGlyTrp 225
         ||| ::: ::::: ||||
Db 1652180 ATCATCTGCACCGACTTCGCAATACGTCAAAAATGGCATGGAAAACTGGATACACGGTTGG 1652239

Qy   226 LysLysAsnGlyTrpLysThrSerAlaGlyLysGluValIleLeuAsnLysGluAspPheVal 245
         ||| ::: ||||
Db 1652240 AAGCGCAACGGCTGGAAAACCGCCTCCAAACAGCCCGTCAAAAACGACGACTTGTGGAAA 1652299

Qy   246 AlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHisValProGlyHisHisSer 265
         ||| ::: :||||
Db 1652300 GAACTGGAACGGCTCTAGTCGGACGGCATCAAGTCAGTTGGGATTGGGTGAAAGGACACGCCG 1652359

Qy   266 GlyPheIleGlyAsnGluGluLeuAsnGluGluAspArgGluAlaArgLeuAlaAlaArgLysGln 283
         ||| :::: |||||
Db 1652360 GGACACGCCGAAAACGAACCGGTGGCAAACCGTGGCCAGCCGCAG 1652413

RESULT 9
US-09-897-516A-658
; Sequence 658, Application US/09897516A
; GENERAL INFORMATION:
; APPLICANT: Corbin, David R.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Huesing, Joseph E.
; APPLICANT: Malvar, Thomas M.
; APPLICANT: Krasomil-Osterfeld, Karina C.
; APPLICANT: Slater, Steven C.
; APPLICANT: Spiridonov, Sergei
; TITLE OF INVENTION: Xenorhabdus sp. Genome Sequences And Uses Thereof
; FILE REFERENCE: 38-21(51847)B
; CURRENT APPLICATION NUMBER: US/09/897,516A
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/215,161
; PRIOR FILING DATE: 2000-06-30
; NUMBER OF SEQ ID NOS: 8415
; SEQ ID NO 658
; LENGTH: 1733
; TYPE: DNA
; ORGANISM: Xenorhabdus sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (874)..(1335)
; OTHER INFORMATION:
US-09-897-516A-658

Alignment Scores:
Pred. No.:   Length: 1733
Score: 1.8e-14   Matches: 56
Percent Similarity: 212.50   Conservative: 17
Best Local Similarity: 47.10%   Mismatches: 71
Query Match: 36.13%   Indels: 11
DB: 13.75%   Gaps: 4
                                5

US-10-054-313-1 (1-286) x US-09-897-516A-658 (1-1733)

Qy   130 ArgAspThrPheSerTyrMetGlyAspPheValValTyrThrAspGlyCysCysSer 149
         ||| :::|||||||||||||||||||

Db   856 AGGAAGAGTCTACCAAATATGAATAAACAGGTAGAAATTTCACCGATGGTTCCTGCCTC 915
Qy   150 SerAsnGlyArgArgLysProArgAlaGlyIleGlyVal--------TyrTrpGlyPro 166
Db   916 GGTAAT--------CCGGGTCCCGGCGGATATGGTGTTTTACTTCGTTATAAACAAAAG 966
Qy   167 GlyHisProLeuAsnValGlyIleArgLeuProGlyArgGlnThrAsnGlnArgAlaGlu 186
Db   967 GAAAAAACCTTAAATAAAGGTTAT--------TTCCGTACAACCAATAACCGGATGGAG 1017
Qy   187 IleHisAlaAlaCysLysLysGlnAlaIleGluGlnAlaLysThrGlnAsnIleAsnLysLeuVal 206
Db 1018 CTAATGGCCGCTATCGCCGCGCTGGAAGCCCTTAAAGAGCCA-----TGCTCAATCATT 1071
Qy   207 LeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsnTrpValGlnGlyTrpLys 226
Db 1072 CTTACCACTGACAGCCAATATGTTCGTCAGGGAATTACCCAGTGGATACAATAACTGGAAA 1131
Qy   227 LysAsnGlyTrpThrSerAlaGlyLysLysGluValIleAsnLysGluAspPheValAla 246
Db 1132 AAACGGGGCTGGAAAAAAGCTGATAAATCCCGGTGATCAATGTTGATCTGTGGCAACGG 1191
Qy   247 LeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHisValProGlyHisSerGly 266
Db 1192 CTGGATAAAGCCATTTCTCCTCACGAAATCGAATGGCATTGGGTCAAGGGGCATACCGGA 1251
Qy   267 PheIleGlyAsnGluGluAlaAspArgLeuAlaArgGluGlyAla 281
Db 1252 CATCGTGAAAACGAATTGTGTGATGAACTCGCCCGTGCAGCGGCT 1296

RESULT 10
US-09-897-516A-659/c
; Sequence 659, Application US/09897516A
; GENERAL INFORMATION:
; APPLICANT: Corbin, David R.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Huesing, Joseph E.
; APPLICANT: Malvar, Thomas M.
; APPLICANT: Krasomil-Osterfeld, Karina C.
; APPLICANT: Slater, Steven C.
; APPLICANT: Spiridonov, Sergei
; TITLE OF INVENTION: Xenorhabdus sp. Genome Sequences And Uses Thereof
; FILE REFERENCE: 38-21(51847)B
; CURRENT APPLICATION NUMBER: US/09/897,516A
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/215,161
; PRIOR FILING DATE: 2000-06-30
; NUMBER OF SEQ ID NOS: 8415
; SEQ ID NO 659
; LENGTH: 1733
; TYPE: DNA
; ORGANISM: Xenorhabdus sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (925)..(1674)
; OTHER INFORMATION:
US-09-897-516A-659

Alignment Scores:
Pred. No.:   1.8e-14   Length: 1733
Score: 212.50   Matches: 56
Percent Similarity: 47.10%   Conservative: 17
Best Local Similarity: 36.13%   Mismatches: 71
Query Match: 13.75%   Indels: 11
DB: Gaps: 4

US-10-054-313-1 (1-286) x US-09-897-516A-659 (1-1733)

Qy   130 ArgAspThrPheSerTyrMetGlyAspPheValValTyrThrAspGlyCysCysSer 149
Db   878 AGGAAGAGTCTACCAAATATGAATAAACAGGTAGAAATTTCACCGATGGTTCCTGCCTC 819
Qy   150 SerAsnGlyArgArgLysProArgAlaGlyIleGlyVal--------TyrTrpGlyPro 166

RESULT 11
US-09-897-516A-660/c
; Sequence 660, Application US/09897516A
; GENERAL INFORMATION:
; APPLICANT: Corbin, David R.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Huesing, Joseph E.
; APPLICANT: Malvar, Thomas M.
; APPLICANT: Krasomil-Osterfeld, Karina C.
; APPLICANT: Slater, Steven C.
; APPLICANT: Spiridonov, Sergei
; TITLE OF INVENTION: Xenorhabdus sp. Genome Sequences And Uses Thereof
; FILE REFERENCE: 38-21(51847)B
; CURRENT APPLICATION NUMBER: US/09/897,516A
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/215,161
; PRIOR FILING DATE: 2000-06-30
; NUMBER OF SEQ ID NOS: 8415
; SEQ ID NO 660
; LENGTH: 1733
; TYPE: DNA
; ORGANISM: Xenorhabdus sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2)..(352)
; OTHER INFORMATION:
US-09-897-516A-660

Alignment Scores:
Pred. No.:   1.8e-14   Length: 1733
Score: 212.50   Matches: 56
Percent Similarity: 47.10%   Conservative: 17
Best Local Similarity: 36.13%   Mismatches: 71
Query Match: 13.75%   Indels: 11
DB: Gaps: 4

US-10-054-313-1 (1-286) x US-09-897-516A-660 (1-1733)

Qy   130 ArgAspThrPheSerTyrMetGlyAspPheValValTyrThrAspGlyCysCysSer 149
Db   878 AGGAAGAGTCTACCAAATATGAATAAACAGGTAGAAATTTCACCGATGGTTCCTGCCTC 819
Qy   150 SerAsnGlyArgArgLysProArgAlaGlyIleGlyVal--------TyrTrpGlyPro 166

Qy   150 SerAsnGlyArgArgLysProArgAlaGlyIleGlyVal--------TyrTrpGlyPro 166
Db   818 GGTAAT--------CCGGGTCCCGGCGGATATGGTGTTTTACTTCGTTATAAACAAAAG 768
Qy   167 GlyHisProLeuAsnValGlyIleArgLeuProGlyArgGlnThrAsnGlnArgAlaGlu 186
Db   767 GAAAAAACCTTAAATAAAGGTTAT--------TTCCGTACAACCAATAACCGGATGGAG 717
Qy   187 IleHisAlaAlaCysLysLysGlnAlaIleGluGlnAlaLysThrGlnAsnIleAsnLysLeuVal 206
Db   716 CTAATGGCCGCTATCGCCGCGCTGGAAGCCCTTAAAGAGCCA-----TGCTCAATCATT 663
Qy   207 LeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsnTrpValGlnGlyTrpLys 226
Db   662 CTTACCACTGACAGCCAATATGTTCGTCAGGGAATTACCCAGTGGATACAATAACTGGAAA 603
Qy   227 LysAsnGlyTrpThrSerAlaGlyLysLysGluValIleAsnLysGluAspPheValAla 246
Db   602 AAACGGGGCTGGAAAAAAGCTGATAAATCCCGGTGATCAATGTTGATCTGTGGCAACGG 543
Qy   247 LeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHisValProGlyHisSerGly 266
Db   542 CTGGATAAAGCCATTTCTCCTCACGAAATCGAATGGCATTGGGTCAAGGGGCATACCGGA 483
Qy   267 PheIleGlyAsnGluGluAlaAspArgLeuAlaArgGluGlyAla 281
Db   482 CATCGTGAAAACGAATTGTGTGATGAACTCGCCCGTGCAGCGGCT 438

Db   818 GGTAAT--------CCGGGTCCGGCGGATATGGTGTTTACTTCGTTATAAACAAAAG 768

Qy   167 GlyHisProLeuAsnValGlyIleArgLeuProGlyArgGlnThrArgAsnGlnArgAlaGlu 186
              |||||||
Db   767 GAAAAAACCTTAAATAAAGGTTAT-------TTCCGTACACCAATAACCGGATGGAG 717

Qy   187 IleHisAlaAlaCysLysAlaIleGluGlnAlaLysThrGlnAsnIleAsnLysLeuVal 206
          ::: |||||||| :::::
Db   716 CTAATGGGCCGCTATCGCCGCCGCTGGAAGCCCTAAAGAGCCA------TGCTCAATCATT 663

Qy   207 LeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsnTrpValGlnGlyTrpLys 226
              ||| :::
Db   662 CTTACCACTGACAGCCAATATGTTCGTCAGGAATTACCCAGTGGAATACATAACTGGAAA 603

Qy   227 LysAsnGlyThrTrpLysTyrSerAlaGlyIleGlyLysGluAsnLysGluAspPheValAla 246
          ||||||| :::   :::
Db   602 AAACGGGGCTGGAAAAAGCTGATAAATCCCCGGTATCAATGTTGATCTGTGGCAACGG 543

Qy   247 LeuGluArgLeuThrPheGlnGlyMetAspIleGlyTrpMetHisValProGlyHisSerGly 266
          |||:::::: ||||||||
Db   542 CTGGATAAAAGCCATTTCTCCTCACGAAATGGCATTGGGTCAAGGGGCATACCGGA 483

Qy   267 PheIleGlyAsnGluGluAlaAspArgLeuAlaArgGluAlaArgGluAlaAla 281
          |||
Db   482 CATCGTGAAAAACGAATTGTGATGAACTCGCCCGTGCAGCCGCT 438

RESULT 12
US-10-684-141-41
; Sequence 41, Application US/10684141
; GENERAL INFORMATION:
; APPLICANT: Furusawa, Mitsuru
; TITLE OF INVENTION: METHOD AND SYSTEM FOR RAPIDLY CONFERRING
; TITLE OF INVENTION: A DESIRED TRAIT TO AN ORGANISM
; FILE REFERENCE: 690116.401
; CURRENT APPLICATION NUMBER: US/10/684,141
; CURRENT FILING DATE: 2003-10-10
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 41
; LENGTH: 1592
; TYPE: DNA
; ORGANISM: Escherichia coli
US-10-684-141-41

Alignment Scores:
Pred. No.:   Length: 1592
Score: 4.71e-14   Matches: 52
Percent Similarity: 208.50   Conservative: 16
Best Local Similarity: 47.89%   Mismatches: 69
Query Match: 36.62%   Indels: 5
DB: 13.49%   Gaps: 2

US-10-054-313-1 (1-286) x US-10-684-141-41 (1-1592)

Qy   140 ValValValTyrThrAspGlyCysCysSerSerAsnGlyArgSerAsnGlyArgArgLysProArgAlaGly 159
          :::::::::|| |||||||
Db   1088 GTAGAAATTTTCACCGATCGTTCCGTCTGGGCAATCCAGGACCTGGGCGGTTACGGCGCT 1147

Qy   160 IleGlyValTyrTrpGlyProGlyHisProLeuProGlyHisProLeuArgLeuProGlyArg 179
          ||| ---CGC
Db   1148 ATTTTACGCTATCGCGGACGCGGAGAAACCTTAGCCGTGGTGGTTACACC------CGC 1198

Qy   180 GlnThrArgGlnArgAlaGlyIleHisAlaAlaCysLysAlaIleGluGlnAlaLysThr 199
          :::::::||
Db   1199 ACCACCAACACCGTATGGAGTTGATGGCGCCATTGTCGCGCTGGAGGCCGTTAAAAGAA 1258

Qy   200 GlnIleAsnIleAsnLysLeuValLeuValTyrThrAspSerMetPheThrIleAsnGlyIleThr 219
          ||| :::
Db   1259 CAT-----TGCGAAGTCATTTTGAGTACCGACAGCCAGCCATGTCCGCCACAGGTATCACC 1312

Qy   220 AsnTrpValGlnGlyTrpLysLysAsnGlyThrTrpLysTyrSerAlaGlyIleGlyLysGluValIle 239
          ||| ::::::
Db   1313 CAGTGGGATCCATAACGTGGCTGCTGGAAAAAACGTGGCTGCTGGAAAAAACCAGTAAAA 1372

Db   240 AsnLysGluAspPheValAlaLeuGlyArgLeuThrGlnGlyMetAspIleGlnTrpMet 259
          ||| ::: .||||||||||||
Db   1373 AATGTCGATCTCGGCAACGTCTTGATCGTCGCATTGGGCGCAGCAGCATCAAATGGGAA 1432

Qy   260 HisValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgGlu 279
          ||||||||||:::|
Db   1433 TGGGTTAAAGGCCATGCCGGACACCCGAAAAACGAACGCTGTGATGAACTGCGTCGTGCC 1492

Qy   280 GlyAla 281
          |||
Db   1493 GCGGCG 1498

RESULT 13
US-10-297-465B-1/c
; Sequence 1, Application US/10297465B
; GENERAL INFORMATION:
; APPLICANT: Simpson, Andrew
; APPLICANT: Reinach, Fernando
; APPLICANT: Setubal, Joao
; APPLICANT: Medianis, Joao
; APPLICANT: Arruda, Paulo
; TITLE OF INVENTION: Isolated Genome of Xylella Fastidiosa and Uses Thereof
; FILE REFERENCE: FAPESP 202 US (10213376)
; CURRENT APPLICATION NUMBER: US/10/297,465B
; CURRENT FILING DATE: 2003-08-25
; PRIOR APPLICATION NUMBER: PCT/IB01/01618
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: 60/209,906
; PRIOR FILING DATE: 2001-06-17
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 2731748
; TYPE: DNA
; ORGANISM: Xylella fastidiosa
US-10-297-465B-1

Alignment Scores:
Pred. No.:   Length: 2731748
Score: 5.84e-07   Matches: 51
Percent Similarity: 194.50   Conservative: 17
Best Local Similarity: 46.90%   Mismatches: 54
Query Match: 35.17%   Indels: 23
DB: 12.58%   Gaps: 4

US-10-054-313-1 (1-286) x US-10-297-465B-1 (1-2731748)

Qy   143 TyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGlyIleGlyVal 162
                                                     ---HisProLeuAsnValGlyIleArgLeu 176
Db 2049117 -----GGTCCTGGTGGTGGTTGGGCTGTGTGTTGCGTTATAAAAACAATGAGAAGAGTTA 2049064

Qy   163 TyrTrpGlyProGly-------   HisProLeuAsnValGlyIleArgLeu 176
           :::
Db 2049063 GTTGGTGGTGGAGCTTGATACGACCAATAATACGGCAATAATGAGCTCGATGGCCGATTATGGCT 2049004

Qy   177 ProGlyArgGln--------   ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaAla 193
           :::
Db 2049003 CTGGAG--------CGGTTGAGTGAGCTTGAGAATTGGATGGCGTTAAAGAAGCCGTTATGGGATGGCT 2048950

Qy   194 IleGluGlnAlaLysThrGlnAsnIleAsnLysLeuAsnLysValLeuValTyrThrAspSerMetPhe 213
                     CGGTTGAGGTGAGTTGAGAATTGGATGGCGTTAAAGAAGCCGTTAGTTCGCAGTAT 2048890
Db 2049003

Qy   214 ThrIleAsnGlyIleThrAsnTrpValGlnGlyTrpLysLysAsnGlyThrTrpLysTyrSer 233
          :::::::
Db 2048949 GTCCGGCCAAGGGATCACTGAGTGGATGTCCCGTTGGGTACGGCGTGGTTGGAAAAACTGCC 2048890

Qy   234 AlaGlyLysGluValIleLysAsnGlyGluAspPheValAlaLeuGluArgLeuThrGlnGly 253
           :::::::::
Db 2048889 GCCGGTGACCCGTCACCGTCAAGAATCGTGATCTATGGGAGGGGTTGTGTGCGCACGCAGCGG 2048830

Qy   254 MetAspIleGlnTrpMetHisValProGlyHisSerGlyPheIleGlyAsnGluGluAla 273

Db   2048829 CACATGGTGAATGGTGTTGGGTCAAGGCTCACAATGTGATTCGGGATAATGAACGGGGTT 2048770
              :::::|||| ||:::||| ||||||
Qy   274 AspArgLeuAlaArg 278
              ||:|| |||||
Db   2048769 GATGTATTGGCCCGC 2048755

RESULT 14
US-10-425-114A-34337
; Sequence 34337, Application US/10425114A
GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114A
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 34337
; LENGTH: 1545
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLMO17253C07_FLI
US-10-425-114A-34337

Alignment Scores:
Pred. No.:   1.22e-09   Length: 1545
Score: 171.50   Matches: 45
Percent Similarity: 49.67%   Conservative: 30
Best Local Similarity: 29.80%   Mismatches: 57
Query Match: 11.09%   Indels: 19
                                             Gaps: 6

US-10-054-313-1 (1-286) x US-10-425-114A-34337 (1-1545)

Qy   140 ValValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGly 159
           :::|||||||||:||:|||:||||||||| ||| |
Db   377 GTAACTCTCTACACCGATGGTGCCTGCTCCATGAAT------------------CCTGGC 418

Qy   160 IleGlyValTyr---------TrpGlyProGly-----HisProLeuAsnVal 172
           |||:||| ||| :||:| ||| ::|::
Db   419 CCTGGTGGCCATGGCGCAGTGATTCTCACGGTGATGGTCGTCGGGAGGAATTGTCCGCC 478

Qy   173 GlyIleArgLeuProGlyGlnArgGlnThrAsnGlnArgAlaGlyIleHisAlaAlaCysLys 192
           |||:||| ||::|:|:| ||||:| |||||||||
Db   479 GGTTATAAAATG------ACCACCAATAATCGCATGGAAATTATGGGGGCGATCGGC 529

Qy   193 AlaIleGluGlnAsnAlaLysThrGlnAsnIleAsnLysLeuValLeuTyrThrAspSerMet 212
           ||:   ::: ||| || :|||::::||:|||:::||::|:||
Db   530 GCCTTG------AGTCATTACAAGAACCGTCCCAAGTGGTTGCTCTATACCGATTCCCGC 583

Qy   213 PheThrIleAsnGlyIleThr---AsnTrpValGlnGlnTrpLysLysAsnGlyTrpLys 231
           :::||| ||| :|| :||::||| :|:::|||:::
Db   584 TATATGGTCGATGCCATGAGCAAGGGTTGGGCAAGAAATGGAAAGCCAACGATGGCAA 643

Qy   232 ThrSerAlaGlyLyLysGluValIleAsnLysGluAspPheValAlaLeuGluArgLeuThr 251
           ::||| ::| ::| :::
Db   644 AGGAATGCCAAGGAAAAGGCCAAAAACCCCGACCTGTGGAAACGATGTTAACGGCTGTGC 703

Qy   252 GlnGlyMetAspIleGlnTrpMetHisValProGlyHisSerGlyPheIleGlyAsnGlu 271
           ::: ||| :::
Db   704 GAAAAACATCAAGTTACATTCCAGTGGGTCAAAGCCACCCTGGCAATAAGGAAAATGAG 763

Qy   272 GluAlaAspArgLeuAlaArgGluGlyAlaLys 282
           ||||:|| :::
Db   764 CGGTGCGATGCCGAATTGCTCAAGGAATCCGG 796

```
                    GenCore version 5.1.6
              Copyright (c) 1993 - 2003  Compugen Ltd.


OM protein - nucleic search, using frame_plus_p2n model

Run on:        December 17, 2003, 19:17:25 ; Search time 3541 Seconds
                                 (without alignments)
                                 2678.577 Million cell updates/sec

Title:         US-10-054-313-1
Perfect score: 1546
Sequence:      1 MSWLLFLAHRVALAALPCRR.........FIGNEEADRLAREGAKQSED 286

Scoring table: BLOSUM62
               Xgapop  10.0 , Xgapext   0.5
               Ygapop  10.0 , Ygapext   0.5
               Fgapop   6.0 , Fgapext   7.0
               Delop    6.0 , Delext    7.0

Searched:      33363688 seqs, 16581889874 residues

Total number of hits satisfying chosen parameters:       66727376

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10054313/runat_17122003_150745_24277/app_query.fasta_1.455
-DB=Pending_Patents_NA_Main -QFMT=fastap -SUFFIX=rnpm -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10054313 @CGN_1_3914 @runat_17122003_150745_24277 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=30 -XGAPOP=10 -XGAPEXT=0.5  -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :     Pending_Patents_NA_Main:*
      1: /cgn2_6/ptcdata/2/pna/PCTUS_COMB.seq:*
      2: /cgn2_6/ptcdata/2/pna/PCTUS_COMB.seq.old:*
      3: /cgn2_6/ptcdata/2/pna/US06_COMB.seq:*
      4: /cgn2_6/ptcdata/2/pna/US07_COMB.seq:*
      5: /cgn2_6/ptcdata/2/pna/US080_COMB.seq:*
      6: /cgn2_6/ptcdata/2/pna/US081_COMB.seq:*
      7: /cgn2_6/ptcdata/2/pna/US082_COMB.seq:*
      8: /cgn2_6/ptcdata/2/pna/US083_COMB.seq:*
      9: /cgn2_6/ptcdata/2/pna/US084_COMB.seq:*
     10: /cgn2_6/ptcdata/2/pna/US085_COMB.seq:*
     11: /cgn2_6/ptcdata/2/pna/US086_COMB.seq:*
     12: /cgn2_6/ptcdata/2/pna/US087_COMB.seq:*
     13: /cgn2_6/ptcdata/2/pna/US088_COMB.seq:*
     14: /cgn2_6/ptcdata/2/pna/US089_COMB.seq:*
     15: /cgn2_6/ptcdata/2/pna/US090_COMB.seq:*
     16: /cgn2_6/ptcdata/2/pna/US091_COMB.seq:*
     17: /cgn2_6/ptcdata/2/pna/US092A_COMB.seq:*
     18: /cgn2_6/ptcdata/2/pna/US092B_COMB.seq:*
     19: /cgn2_6/ptcdata/2/pna/US093A_COMB.seq:*
     20: /cgn2_6/ptcdata/2/pna/US093B_COMB.seq:*
     21: /cgn2_6/ptcdata/2/pna/US094_COMB.seq:*
     22: /cgn2_6/ptcdata/2/pna/US095A_COMB.seq:*
     23: /cgn2_6/ptcdata/2/pna/US095B_COMB.seq:*
     24: /cgn2_6/ptcdata/2/pna/US095C_COMB.seq:*
     25: /cgn2_6/ptcdata/2/pna/US095D_COMB.seq:*
     26: /cgn2_6/ptcdata/2/pna/US096A_COMB.seq:*
     27: /cgn2_6/ptcdata/2/pna/US096B_COMB.seq:*
     28: /cgn2_6/ptcdata/2/pna/US096C_COMB.seq:*
     29: /cgn2_6/ptcdata/2/pna/US096D_COMB.seq:*
     30: /cgn2_6/ptcdata/2/pna/US096E_COMB.seq:*
     31: /cgn2_6/ptcdata/2/pna/US097A_COMB.seq:*
     32: /cgn2_6/ptcdata/2/pna/US097B_COMB.seq:*
     33: /cgn2_6/ptcdata/2/pna/US097C_COMB.seq:*
     34: /cgn2_6/ptcdata/2/pna/US098A_COMB.seq:*
     35: /cgn2_6/ptcdata/2/pna/US098B_COMB.seq:*
     36: /cgn2_6/ptcdata/2/pna/US098C_COMB.seq:*
     37: /cgn2_6/ptcdata/2/pna/US098D_COMB.seq:*
     38: /cgn2_6/ptcdata/2/pna/US099A_COMB.seq:*
     39: /cgn2_6/ptcdata/2/pna/US099B_COMB.seq:*
     40: /cgn2_6/ptcdata/2/pna/US099C_COMB.seq:*
     41: /cgn2_6/ptcdata/2/pna/US099D_COMB.seq:*
     42: /cgn2_6/ptcdata/2/pna/US099E_COMB.seq:*
     43: /cgn2_6/ptcdata/2/pna/US099F_COMB.seq:*
     44: /cgn2_6/ptcdata/2/pna/US100A_COMB.seq:*
     45: /cgn2_6/ptcdata/2/pna/US100B_COMB.seq:*
     46: /cgn2_6/ptcdata/2/pna/US101A_COMB.seq:*
     47: /cgn2_6/ptcdata/2/pna/US101B_COMB.seq:*
     48: /cgn2_6/ptcdata/2/pna/US102A_COMB.seq:*
     49: /cgn2_6/ptcdata/2/pna/US102B_COMB.seq:*
     50: /cgn2_6/ptcdata/2/pna/US103A_COMB.seq:*
     51: /cgn2_6/ptcdata/2/pna/US103B_COMB.seq:*
     52: /cgn2_6/ptcdata/2/pna/US104A_COMB.seq:*
     53: /cgn2_6/ptcdata/2/pna/US104B_COMB.seq:*
     54: /cgn2_6/ptcdata/2/pna/US6000_COMB.seq:*
     55: /cgn2_6/ptcdata/2/pna/US6001_COMB.seq:*
     56: /cgn2_6/ptcdata/2/pna/US6002_COMB.seq:*
     57: /cgn2_6/ptcdata/2/pna/US6003_COMB.seq:*
     58: /cgn2_6/ptcdata/2/pna/US6004_COMB.seq:*
     59: /cgn2_6/ptcdata/2/pna/US6005_COMB.seq:*
     60: /cgn2_6/ptcdata/2/pna/US6006_COMB.seq:*
     61: /cgn2_6/ptcdata/2/pna/US6007_COMB.seq:*
     62: /cgn2_6/ptcdata/2/pna/US6008_COMB.seq:*
     63: /cgn2_6/ptcdata/2/pna/US6009_COMB.seq:*
     64: /cgn2_6/ptcdata/2/pna/US6010_COMB.seq:*
     65: /cgn2_6/ptcdata/2/pna/US6011_COMB.seq:*
     66: /cgn2_6/ptcdata/2/pna/US6012_COMB.seq:*
     67: /cgn2_6/ptcdata/2/pna/US6013_COMB.seq:*
     68: /cgn2_6/ptcdata/2/pna/US6014_COMB.seq:*
     69: /cgn2_6/ptcdata/2/pna/US6015_COMB.seq:*
     70: /cgn2_6/ptcdata/2/pna/US6016_COMB.seq:*
     71: /cgn2_6/ptcdata/2/pna/US6017_COMB.seq:*
     72: /cgn2_6/ptcdata/2/pna/US6018_COMB.seq:*
     73: /cgn2_6/ptcdata/2/pna/US6019_COMB.seq:*
     74: /cgn2_6/ptcdata/2/pna/US6020_COMB.seq:*
     75: /cgn2_6/ptcdata/2/pna/US6021_COMB.seq:*
     76: /cgn2_6/ptcdata/2/pna/US6022_COMB.seq:*
     77: /cgn2_6/ptcdata/2/pna/US6023A_COMB.seq:*
     78: /cgn2_6/ptcdata/2/pna/US6023B_COMB.seq:*
     79: /cgn2_6/ptcdata/2/pna/US6023C_COMB.seq:*
     80: /cgn2_6/ptcdata/2/pna/US6024_COMB.seq:*
     81: /cgn2_6/ptcdata/2/pna/US6025_COMB.seq:*
     82: /cgn2_6/ptcdata/2/pna/US6026_COMB.seq:*
     83: /cgn2_6/ptcdata/2/pna/US6027_COMB.seq:*
     84: /cgn2_6/ptcdata/2/pna/US6029_COMB.seq:*
     85: /cgn2_6/ptcdata/2/pna/US6030_COMB.seq:*
     86: /cgn2_6/ptcdata/2/pna/US6031_COMB.seq:*
     87: /cgn2_6/ptcdata/2/pna/US6032_COMB.seq:*
     88: /cgn2_6/ptcdata/2/pna/US6033_COMB.seq:*
     89: /cgn2_6/ptcdata/2/pna/US6034_COMB.seq:*
     90: /cgn2_6/ptcdata/2/pna/US6035_COMB.seq:*
     91: /cgn2_6/ptcdata/2/pna/US6036_COMB.seq:*
     92: /cgn2_6/ptcdata/2/pna/US6037_COMB.seq:*
     93: /cgn2_6/ptcdata/2/pna/US6038_COMB.seq:*
     94: /cgn2_6/ptcdata/2/pna/US6039_COMB.seq:*
     95: /cgn2_6/ptcdata/2/pna/US6040_COMB.seq:*
     96: /cgn2_6/ptcdata/2/pna/US6041_COMB.seq:*
     97: /cgn2_6/ptcdata/2/pna/US6042_COMB.seq:*
     98: /cgn2_6/ptcdata/2/pna/US6043_COMB.seq:*
     99: /cgn2_6/ptcdata/2/pna/US6044_COMB.seq:*
    100: /cgn2_6/ptcdata/2/pna/US6045_COMB.seq:*
    101: /cgn2_6/ptcdata/2/pna/US6046_COMB.seq:*
```

```
102: /cgn2_6/ptodata/2/pna/US6047_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                                    SUMMARIES

                    %
Result       Query
No.   Score  Match  Length  DB  ID                        Description
----  -----  -----  ------  --  ----------------------    -----------
1     1526   98.7   1652    30  US-09-698-010-12700       Sequence 12700, A
2     1526   98.7   1652    30  US-09-698-013-6258        Sequence 6258, Ap
3     1526   98.7   1652    30  US-09-699-997-12450       Sequence 12450, A
4     1526   98.7   1652    31  US-09-700-000-5074        Sequence 5074, Ap
5     1526   98.7   1807    47  US-10-198-846-12889       Sequence 12889, A
6     1526   98.7   2124    33  US-09-770-171-1164        Sequence 1164, Ap
7     1526   98.7   2129    33  US-09-785-276A-30281      Sequence 30281, A
8     1526   98.7   2129    51  US-10-357-930-30281       Sequence 30281, A
9     1523   98.5   1150    33  US-09-763-233-41          Sequence 41, Appl
10    1523   98.5   1150    65  US-60-115-639-32          Sequence 32, Appl
11    1522   98.4    965     1  PCT-US02-31373-85         Sequence 85, Appl
12    1522   98.4    965    49  US-10-262-511-85          Sequence 85, Appl
13    1522   98.4    965    49  US-10-262-511A-85         Sequence 85, Appl
14    1522   98.4   1147     1  PCT-US02-18947-841        Sequence 841, App
15    1522   98.4   1147    47  US-10-172-118-841         Sequence 841, App
16    1522   98.4   1147    50  US-10-342-887-841         Sequence 841, App
17    1510   97.7   1629    71  US-60-172-373-7744        Sequence 7744, Ap
18    1510   97.7   1634    82  US-60-278-258-8522        Sequence 8522, Ap
19    1508   97.5   1015    31  US-09-705-256A-3471       Sequence 3471, Ap
20    1508   97.5   1015    70  US-60-164-285-3471        Sequence 3471, Ap
21    1504   97.3   1135    20  US-09-359-922-4578        Sequence 4578, Ap
22    1504   97.3   1135    20  US-09-359-922-4578        Sequence 4578, Ap
23    1504   97.3   1135    38  US-09-919-002-4578        Sequence 4578, Ap
24    1447.5 93.6   1533    47  US-10-170-235-15341       Sequence 15341, A
25    1442.5 93.3   1533   100  US-60-452-680-3892        Sequence 3892, Ap
26    1218   78.8   1533     1  PCT-US02-14199-49         Sequence 49, Appl
27    1218   78.8    979    46  US-10-136-826A-49         Sequence 49, Appl
28    1189.5 76.9   1417    46  US-10-144-771-9913        Sequence 9913, Ap
29    1189.5 76.9   1417    91  US-60-360-207-9913        Sequence 9913, Ap
30    1174   75.4   4741    40  US-09-948-128-112         Sequence 112, App
31    1150.5 74.4   5508    40  US-09-948-128-112         Sequence 113, App
32    1045.5 67.6    631     1  PCT-US02-31373-79         Sequence 79, Appl
33    1045.5 67.6    631    49  US-10-262-511-79          Sequence 79, Appl
34    1045.5 67.6    631    49  US-10-262-511A-79         Sequence 79, Appl
35    1045.5 67.6    716     1  PCT-US02-31373-77         Sequence 77, Appl
36    1045.5 67.6    716    49  US-10-262-511-77          Sequence 77, Appl
37    1045.5 67.6    716    49  US-10-262-511A-77         Sequence 77, Appl
38    1010   65.3  387372  101  US-60-465-241-51783       Sequence 51783, A
39    979.5  63.4    586     1  PCT-US02-31373-81         Sequence 81, Appl
40    979.5  63.4    586    49  US-10-262-511-81          Sequence 81, Appl
41    979.5  63.4    586    49  US-10-262-511A-81         Sequence 81, Appl
42    926    59.9    550    23  US-09-534-855-365         Sequence 365, App
43    905    58.5    545    33  US-09-785-276A-45806      Sequence 45806, A
44    905    58.5    545    51  US-10-357-930-45806       Sequence 45806, A
45    895    57.9   1248    47  US-10-170-235-15216       Sequence 15216, A
```

```
                                   ALIGNMENTS

RESULT 1
US-09-698-010-12700
; Sequence 12700, Application US/09698010
; GENERAL INFORMATION:
; APPLICANT: Williamson, Mark
; APPLICANT: Shyjan, Andrew W.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 1600.2029-001
; CURRENT APPLICATION NUMBER: US/09/698,010
; CURRENT FILING DATE:  2000-10-27
; PRIOR APPLICATION NUMBER: 60/162,358
; PRIOR FILING DATE: 1999-10-29
```

```
; NUMBER OF SEQ ID NOS: 15684
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12700
;   LENGTH: 1652
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-698-010-12700

Alignment Scores:
Pred. No.:                1.33e-119
Score:                    1526.00
Percent Similarity:       99.30%
Best Local Similarity:    98.60%
Query Match:              98.71%
DB:                       30

                   Length:         1652
                   Matches:        282
                   Conservative:     2
                   Mismatches:       2
                   Indels:           0
                   Gaps:             0

US-10-054-313-1 (1-286) x US-09-698-010-12700 (1-1652)

Qy     1  MetSerTrpLeuPheLeuAlaHisArgValAlaAlaAlaAlaLeuProCysArgArg         20
Db   112  ATGAGCTGGCTTCTGTTCCTGGCCCACAGAGTCGCCTTGCCTTGCCTGCCCTGCCGCCGC    171

Qy    21  GlySerArgGlyPheGlyMetPheTyrAlaValArgArgGlyArgGlyArgGlyValPhe       40
Db   172  GGCTCTCGCGGGGTTCGGGATGTTCTATGCCGTGAGGAGGGGCCGCAAGAGACCGGGGTCTTT    231

Qy    41  LeuThrTrpAsnGluCysArgAlaGlnValAspArgArgPheProAlaAlaArgPheLysLys    60
Db   232  CTGACCTGGAATGAGTGCAGAGCACAGGTGGACCCGGTTCCTGCTGCCAGATTTAAGAAG     291

Qy    61  PheAlaThrGluAspGluAlaTrpProAlaPheValArgLysSerAlaSerProGluValSer    80
Db   292  TTTGCCACAGAGGATGAGGCCTGGCCTTTGCAGGAAATCGCAAGCCCGGAAGTTTCA       351

Qy    81  GluGlyHisGluAsnGlnHisGlyTyrGlnGluSerGluAlaLysProGlyGlyLysArgLeuArg  100
Db   352  GAAGGGCATGAAAAATCAACATGGACAAGAATCGGAGGGCGAAAGCCAAGCCAGCGACTCCGT   411

Qy   101  GluProLeuAspGlyValHisGluSerAlaGlnProTyrAlaLysHisMetLysPro        120
Db   412  GAGCCACTGGATGGAATGGACATGAAAGCGCCAGAGCCGTATGCAAAGCACATGAAGCCG     471

Qy   121  SerValGluProProValSerArgAspThrPheSerTyrMetGlyAspPheVal          140
Db   472  AGCGTGGAGCCCGCCGTCTCCAGTAGCAGAGACACGTTTCTACATGGGAGACTTCGTC      531

Qy   141  ValValTyrThrAspGlyTyrCysCysSerSerAsnGlyArgArgLysProArgAlaGlyIle    160
Db   532  GTCGTCACACTGATGGCTGTGCTCCAGTAATGGGCGTGAGAGGCCGCGAGCCAGGAATC     591

Qy   161  GlyValTyrTrpProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGln         180
Db   592  GGCGTTTACTGGGGCCAGGCCATCCTTAAATGTAGGCATTAGGACTTCCTGGGCGCAG      651

Qy   181  ThrAsnGlnArgAlaGluHisAlaAlaCysLysAlaIleGluGluGlnLysAlaThrGln      200
Db   652  ACAAACCAAGACGGGAAATCATGCAGCCTGCAAAGCCATTGACAAGCAAAGACTCAA       711

Qy   201  AsnIleAsnLysLysValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn      220
Db   712  AACATCAATAAACTGGTTCTGTATACACAGACAGTATGTTTACGATAAATGGTATAACTAAC   771

Qy   221  TrpValGlnGlyTrpLysLysGlyThrSerAlaGlyValTrpLysGluValIleAsn         240
Db   772  TGGGTTCAAGGTTGGAAGAAAAATGGGTGGAAGACAAGTGCAGGGAAAGAGTGATCAAC     831

Qy   241  LysGluAspPheValAlaAlaGluLeuArgLeuThrHisGlnGlyTrpAspMetHis         260
Db   832  AAAGAGGACTTTGTGGCCACTGAGGAGAGCTTACCCAGGGGATGGGATGCATTCAGTGGATGCAT    891

Qy   261  ValProGlyHisSerGlyPheLeuGlyAsnGluAlaAspArgGluLeuAlaArgGluGly       280
Db   892  GTTCCTGGTCATTCGGGATTTATAGGCAATGAGAAGCTGACAGATTAGCCAGAGAGGA      951
```

Qy   281 AlaLysGlnSerGluAsp 286
           |||||||||||||||||
Db   952 GCTAAACAATCGGAAGAC 969

RESULT 2
US-09-698-013-6258
; Sequence 6258, Application US/09698013
; GENERAL INFORMATION:
;   APPLICANT: Gearing, David P.
;   APPLICANT: Conrack, Christopher
;   APPLICANT: Kingsbury, Gillian A.
;   APPLICANT: Holtzman, Douglas A.
;   TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
;   TITLE OF INVENTION: THEREFOR
;   FILE REFERENCE: 1600.2013-001
;   CURRENT APPLICATION NUMBER: US/09/698,013
;   CURRENT FILING DATE: 2000-10-27
;   PRIOR APPLICATION NUMBER: 60/152,360
;   PRIOR FILING DATE: 1999-10-29
;   NUMBER OF SEQ ID NOS: 7935
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6258
;   LENGTH: 1652
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-698-013-6258

Alignment Scores:
Pred. No.:
Score: 1.33e-119   Length: 1652
Percent Similarity: 99.30%   Matches: 282
Best Local Similarity: 98.60%   Conservative: 2
Query Match: 98.71%   Mismatches: 2
DB: Indels: 0
                                         Gaps: 0

US-10-054-313-1 (1-286) x US-09-698-013-6258 (1-1652)

Qy   1 MetSerTrpLeuPheLeuAlaHisArgValAlaLeuAlaLeuAlaLeuProCysArgArg 20
Db   112 ATGAGCTGGCTTCTGTTCCTGGCCCACAGAGTCGCCTTGGCCCTGGCCCTGCCGCCGCGC 171

Qy   21 GlySerArgGlyPheGlyMetPheTyrAlaValArgArgGlyArgGlyThrGlyValPhe 40
Db   172 GGCTCTCGCGGGTTCGGGATGTTCTATGCCGTGAGGAGGGGCCGCAAGACCGGGGTCTTT 231

Qy   41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgArgPheProAlaAlaArgPheLysLys 60
Db   232 CTGACCTGGAATGAGTGCAGAGCACAGGTGGACCGGTTCCTGCTGCCAGATTTAAGAAG 291

Qy   61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80
Db   292 TTTGCCACAGAGGATGAGGCCTGGGCCTTTGTCAGGAAATCTGCAAGCCCGGAAGTTTCA 351

Qy   81 GluGlyHisGlnAsnGlnHisGlyGlnLeuSerGluAlaLysProGlyArgLeuArg 100
Db   352 GAAGGGCATGAAAATCAACATGGACAAGGATCAAGGAGGCGAAAGCCAGCAGACTCCGT 411

Qy   101 GluProLeuAspGlyAspHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120
Db   412 GAGCCACTGGATGGAGATGACAAAGGCCAGACGTATGCAAAGCACATGAAGCCG 471

Qy   121 SerValGluProAlaProProValSerProAlaProThrMetGlyAspPheVal 140
Db   472 AGCGTTGGAGCCCGCCGGCCTCCAGTTAGCAGACACGTTTCCCACATGGGAGACTTCGTC 531

Qy   141 ValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgGlyProAlaGlyIle 160
Db   532 GTCGTCTCCACTGATGGCTGCTGCTCCAGTAATGGGCGTAGAAGGCCCGGAGCCAGGAATC 591

Qy   161 GlyValTyrTrpGlyProGlyLeuAsnValGlyIleArgLeuProGlyGlyGln 180
Db   592 GGGGTTTACTGGGGCCAGGCCATCCTTTAAATGTAGGCATTAGACTTCCTGGGCGGGCAG 651

Qy   181 ThrAsnGlnArgAlaGlyIleHisAlaAlaCysLysAlaAlaIleGluArgGluAlaAlaLysAlaLysThrGln 200
Db   652 ACAAACCAAAGAGGGAAATTCATGCAGCCTGCAAGCCTGACAACAAGACAAGACTCA 711

Qy   201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220
Db   712 AACATCAATAAACTGGTTCTGTATACAGACAGTATGTTACGATAAATGTTAACTAAC 771

Qy   221 TrpValGlnGlyTyrTrpLysLysAsnGlyTyrTrpLysSerAlaGlyLysGluValIleAsn 240
Db   772 TGGGTTCAAGGTTGGAAGAAAAATGGGTGGAACAAGTGCAGGGAAAGAGGTGATCAAC 831

Qy   241 LysGluAspPheValAlaLeuGlnLeuArgLeuThrGlnGlyMetAspIleGlnTrpMetHis 260
Db   832 AAAGAGGACTTTGTGGCACTGGAGGAGCTTACCCAGGGATGGACATTCAGTGGATGCAT 891

Qy   261 ValProGlyHisSerGlyPheIleGlyAsnGlnGluAlaAspArgLeuAlaArgGlyGly 280
Db   892 GTTCCTGGTCATTCGGGATTTATAGGCAATGAAGAAGCTGACAGATTAGCCAGAAGAAGGA 951

Qy   281 AlaLysGlnSerGluAsp 286
Db   952 GCTAAACAATCGGAAGAC 969

RESULT 3
US-09-699-997-12450
; Sequence 12450, Application US/09699997
; GENERAL INFORMATION:
;   APPLICANT: Curtis, Rory A.J.
;   APPLICANT: Silos-Santiago, Inmaculada
;   APPLICANT: DiStefano, Peter
;   TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
;   TITLE OF INVENTION: THEREFOR
;   FILE REFERENCE: 1600.2020-001
;   CURRENT APPLICATION NUMBER: US/09/699,997
;   CURRENT FILING DATE: 2000-10-30
;   PRIOR APPLICATION NUMBER: 60/162,359
;   PRIOR FILING DATE: 1999-10-29
;   NUMBER OF SEQ ID NOS: 12714
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12450
;   LENGTH: 1652
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-699-997-12450

Alignment Scores:
Pred. No.:
Score: 1.33e-119   Length: 1652
Percent Similarity: 99.30%   Matches: 282
Best Local Similarity: 98.60%   Conservative: 2
Query Match: 98.71%   Mismatches: 2
DB: Indels: 0
                                         Gaps: 0

US-10-054-313-1 (1-286) x US-09-699-997-12450 (1-1652)

Qy   1 MetSerTrpLeuPheLeuAlaHisArgValAlaLeuAlaLeuAlaLeuProCysArgArg 20
Db   112 ATGAGCTGGCTTCTGTTCCTGGCCCACAGAGTCGCCTTGGCCCTGGCCCTGCCGCCGCGC 171

Qy   21 GlySerArgGlyPheGlyMetPheTyrAlaValArgArgGlyArgGlyThrGlyValPhe 40
Db   172 GGCTCTCGCGGGTTCGGGATGTTCTATGCCGTGAGGAGGGGCCGCAAGACCGGGGTCTTT 231

Qy   41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgArgPheProAlaAlaArgPheLys 60
Db   232 CTGACCTGGAATGAGTGCAGAGCACAGGTGGACCGGTTCCTGCTGCCAGATTTAAGAAG 291

Qy   61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80
Db   292 TTTGCCACAGAGGATGAGGCCTGGGCCTTTGTCAGGAAATCTGCAAGCCCGGAAGTTTCA 351

US-10-054-313-1 (1-286) x US-09-700-000-5074 (1-1652)

```
Qy      1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaAlaLeuProCysArgArg  20
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    112 ATGAGCTGGCTCCTTGTTCCTGGCCCACAGAGTCGCCCTTGGCCGCCCTGCCGCGCCGC 171

Qy     21 GlySerArgGlyPheGlyMetPheTyrAlaValArgArgGlyThrGlyLysHisGlyValPhe  40
          ||||||||||||||||||||||||:||:||||||||||||||||||||||||||||||||
Db    172 GGCTCTCGCGGGTTCGGGATGTTCTATGCCGTGAGGAGGGGCGCAAGACCGGGTCTTT 231

Qy     41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys  60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    232 CTGACCTGGAATGAGTGCAGAGCACAGGTGGACCGGTTTCCTGCTGCCAGATTTAAGAAG 291

Qy     61 PheAlaThrGlyAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer  80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    292 TTTGCCACAGGAGATGAGGCCTGGGCCTTTGTCAGGAAGTCAGCCTCCCCAGAGGTTTCA 351

Qy     81 GluGlyHisGluAsnGlnHisGlyGlnSerGluAlaLysProGlyLysArgLeuArg 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    352 GAAGGGCATGAAAATCAACATGGACAAGAATCGGAGGCAAAGCCAGCAAGGCGACTCCGT 411

Qy    101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120
          ||||||||||||||||||||||||||||||||||||||:::|||||||||||||||||
Db    412 GAGCCACTGGATGGAGATGGACATGCAGAGCCGTATGCAAAGCACATGAAGCCG 471

Qy    121 SerValGluProAlaProProValSerArgArgAspThrPheSerTyrMetGlyAspPheVal 140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    472 AGCGTGGAGCCGGCGCCCTCAGTTAGCAGAGACACGTTTTCCTACATGGGAGACTTCGTC 531

Qy    141 ValValTyrThrAspGlyCysCysSerSerAlaGlyTyrArgArgGlyProArgAlaGlyIle 160
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    532 GTCGTTCACACTGATGGCTGCTGCTCCAGTAATGGGCGTAGAAGGCCGGAGCAGGAATC 591

Qy    161 GlyValTyrProGlyHisProLeuAsnValGlyIleArgLeuProGlyLeuTyrArgGln 180
          ||||||||||||||||||||||||||||||||||||||::|||
Db    592 GGGGTTTACTGGGGCCAGGCCATCCTTAAATGTAGGCATTAGGCCGGAGCAGGGCGCAG 651

Qy    181 ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysValAlaIleGluGlnAlaLysThrGln 200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    652 ACAAACCAAAGAGCGGAAATTCATGCACCTGCAAAGCCATTGAACACAGCAAAGACTTCAA 711

Qy    201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    712 AACATCAATAAACTGGTTCTGTATACACAGACAGTATGTTTACCATAAATGGTATAACTAAC 771

Qy    221 TrpValGlnGlyTrpLysAsnGlyTrpLysThrSerAlaGlyLysGluValIleLysAsn 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    772 TGGGTTCAAGGTTGGAAGAAAAATGGGTGGAAGACAAGTGCAGGGAAGAGAGGTGATCAAC 831

Qy    241 LysGluAspPheValAlaLeuGluThrGlnLeuArgGlyMetAspIleGlnTrpMetHis 260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    832 AAAGAGGACTTTGTGGCACTGGAGAGGCTTACCCAGGGGATGGACATTCAGTGCGATGCAT 891

Qy    261 ValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgGluGlyGly 280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    892 GTTCCTGGTCATTCGGATTTATAGGCAATGAAGAAGCTGACAGATTAGCCAGAAGGA 951

Qy    281 AlaLysGlnSerGluAsp 286
          ||||||||||||||||||
Db    952 GCTAAACAATCGGAAGAC 969
```

RESULT 5
US-10-198-846-12889
; Sequence 12889, Application US/10198846
; GENERAL INFORMATION:
  APPLICANT: Lillie, James
  APPLICANT: Xu, Yongyao
  APPLICANT: Wang, Youzhen
  APPLICANT: Steinmann, Kathleen
  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
  TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
  TITLE OF INVENTION: THERAPY OF BREAST CANCER

```
Qy     81 GluGlyHisGluAsnGlnHisGlyGlnSerGluAlaLysProGlyLysArgLeuArg 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    352 GAAGGGCATGAAAATCAACATGGACAAGAATCGGAGGCAAAGCCAGCAAGGCGACTCCGT 411

Qy    101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120
          ||||||||||||||||||||||||||||||||||||||:::||||||||||||
Db    412 GAGCCACTGGATGGAGATGGACATGCAGAGCCGTATGCAAAGCACATGAAGCCG 471

Qy    121 SerValGluProAlaProProValSerArgArgAspThrPheSerTyrMetGlyAspPheVal 140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    472 AGCGTGGAGCCGGCGCCCTCAGTTAGCAGAGACACGTTTTCCTACATGGGAGACTTCGTC 531

Qy    141 ValValTyrThrAspGlyCysCysSerSerAlaGlyTyrArgArgGlyProArgAlaGlyIle 160
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    532 GTCGTTCACACTGATGGCTGCTGCTCCAGTAATGGGCGTAGAAGGCCGGAGCAGGAATC 591

Qy    161 GlyValTyrProGlyHisProLeuAsnValGlyIleArgLeuProGlyLeuTyrArgGln 180
          ||||||||||||||||||||||||||||||||||||::::||
Db    592 GGGGTTTACTGGGGCCAGGCCATCCTTAAATGTAGGCATTAGGCCGGAGCAGGGCGCAG 651

Qy    181 ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysValAlaIleGluGlnAlaLysThrGln 200
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    652 ACAAACCAAAGAGCGGAAATTCATGCACCTGCAAAGCCATTGAACACAGCAAAGACTTCAA 711

Qy    201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    712 AACATCAATAAACTGGTTCTGTATACACAGACAGTATGTTTACCATAAATGGTATAACTAAC 771

Qy    221 TrpValGlnGlyTrpLysAsnGlyTrpLysThrSerAlaGlyLysGluValIleLysAsn 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    772 TGGGTTCAAGGTTGGAAGAAAAATGGGTGGAAGACAAGTGCAGGGAAGAGAGGTGATCAAC 831

Qy    241 LysGluAspPheValAlaLeuGluThrGlnLeuArgGlyMetAspIleGlnTrpMetHis 260
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db    832 AAAGAGGACTTTGTGGCACTGGAGAGGCTTACCCAGGGGATGGACATTCAGTGCGATGCAT 891

Qy    261 ValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgGluGlyGly 280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    892 GTTCCTGGTCATTCGGATTTATAGGCAATGAAGAAGCTGACAGATTAGCCAGAAGGA 951

Qy    281 AlaLysGlnSerGluAsp 286
          ||||||||||||||||||
Db    952 GCTAAACAATCGGAAGAC 969
```

RESULT 4
US-09-700-000-5074
; Sequence 5074, Application US/09700000
; GENERAL INFORMATION:
  APPLICANT: Richardson, Jennifer
  APPLICANT: Shyjan, Andrew W.
  TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
  TITLE OF INVENTION: THEREFOR
  FILE REFERENCE: 1600.2022-001
  CURRENT APPLICATION NUMBER: US/09/700,000
  CURRENT FILING DATE: 2000-10-30
  PRIOR APPLICATION NUMBER: 60/162,619
  PRIOR FILING DATE: 1999-10-29
  NUMBER OF SEQ ID NOS: 7171
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5074
; LENGTH: 1652
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-700-000-5074

Alignment Scores:
Pred. No.:
Score: 1.33e-119   Length: 1652
Percent Similarity: 99.30%   1526.00   Matches: 282
Best Local Similarity: 98.60%   Conservative: 2
Query Match: 98.71%   Mismatches: 2
DB: Indels: 0
                                    Gaps: 0

```
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-19
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12889
; LENGTH: 1807
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1802, 1803, 1804, 1805, 1806, 1807
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-12889

Alignment Scores:
Pred. No.:
Score:                  1.47e-119
Percent Similarity:     1526.00
Best Local Similarity:  99.30%
Query Match:            98.60%
DB:                     98.71%
                        47
```

```
Length:        1807
Matches:       282
Conservative:  2
Mismatches:    2
Indels:        0
Gaps:          0
```

```
US-10-054-313-1 (1-286) x US-10-198-846-12889 (1-1807)

Qy    1  MetSerTrpLeuLeuPheLeuLeuAlaHisArgValAlaLeuAlaAlaLeuAlaAlaLeuProCysArgArg   20
Db  112  ATGAGCTGGCTTCTGTTCCTGCTCCTGGCCCACAGAGTCGCCTTGGCCGCCCTTGCCCTGCCCGCGCCGC  171

Qy   21  GlySerArgGlyPheGlyMetPheTyrAlaValArgArgGlyArgLysThrGlyValPhe   40
Db  172  GGCTCTCGCGGGTTCGGGATGTTCTATGCCGTGAGGAGGGGCCGCAAGACCGGGGTCTTT  231

Qy   41  LeuThrTrpAsnGluCysArgAlaGlnValAlaAspArgArgPheProAlaAlaArgPheLysLys   60
Db  232  CTGACCTGGAATGAGTGCAGAGCACAGGTGGACCGTTTCCTGCTGCCAGATTTAAGAAG  291

Qy   61  PheAlaThrGluAspArgAlaTrpAlaPheValArgLysSerAlaSerProGluValSer   80
Db  292  TTTGCCACAGAGGATGAGGCCTGGGCCTTTGTCAGGAAAATCTGCAAGCCCGGAAGTTCA  351

Qy   81  GluGlyHisGluAsnGlnHisGlyGlnGluSerGluLeuGluProGlyLysLysArgLeuArg  100
Db  352  GAAGGGCATGAAAATCAACATGACAAGGCCGAGGCGGAAAGCGGAAGCGGACTCCGT  411

Qy  101  GluProLeuAspGlyAspGlyHisGluSerAlaGlnIleProTyrAlaLysHisMetLysPro  120
Db  412  GAGCCACTGGATGGAGATGGACATGAGAAAGCGCAGACCTGCAAAGCCATGAAGCCG  471

Qy  121  SerValGluProAlaProProValSerArgAspAspArgAspThrPheSerTyrMetGlyAspPheVal  140
Db  472  AGCGTGGAGCCGCCGCCCGTCTCCAGGGACGACAGGGACACGTTTTCCTACATGGGAGACTTCGTC  531

Qy  141  ValValTyrThrAspGlyTyrTrpLysAsnGlyTyrThrSerArgAlaGlyTyrThrSerArgAlaAsn  160
Db  532  GTCGTCTACACTGATGGCTGCTGCTCCAGTAATGGGCCTAGAAGGCCCGAGCCAGGAATC  591

Qy  161  GlyValTyrTrpGlyProLeuAsnValGlyIleAsnGlyGlyLeuProGlyLysArgGln  180
Db  592  GGGGTTTACTGGGGCCCAGGGCCAATCCTTAAATGTAGCATTAGCGGAAGGCCTGCGGCAG  651

Qy  181  ThrAsnGlnArgAlaGlyGluIleIleHisAlaAlaCysLysLysAlaIleGluGlnAlaLysThrGln  200
Db  652  ACAAACAAAGAGCGAGGCGGAAATTCATGCAGCCTGCAAAGCCATTGAACAAGCAAAGACTCAA  711

Qy  201  AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn  220
Db  712  AACTCAATAAAACTGGTTCTGTATACAGAGTATGTTACCGATAAATGGTATAACTAAC  771

Qy  221  TrpValGlnGlyTyrTrpLysSerAsnGlyTyrThrSerArgAlaGlyTyrThrSerArgAlaGlyValValGluValIleAsn  240
Db       GTCGTC...
```

```
Db  772  TGGGTTCAAGTTGGAAGAAAAATGGTGGAAGACAAGTGCAGGAAGAGGTGATCAAC  831
Qy  241  LysGluAspPheValAlaLeuGlyArgLeuThrGlnGlyMetAspIleGlnTrpMetHis  260
Db  832  AAAGAGGACTTTGTGCCACTGGAGAGGCTTACCCAGGGATGGACATTCAGTGATGCAT  891
Qy  261  ValProGlyHisSerGlyPheIleGlyLysAspArgGluAlaAspArgLeuAlaArgGluGly  280
Db  892  GTTCCTGGTCATTCGGGATTTATAGGCAATGAATGAAGAGCTGACAGATTAGCCAGAGAAGGA  951
Qy  281  AlaLysGlnSerGluAsp  286
Db  952  GCTAAACAATCGGAAGAC  969
```

```
RESULT 6
US-09-770-171-1164
; Sequence 1164, Application US/09770171
; GENERAL INFORMATION:
; APPLICANT: Shyjan, Andrew W.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 1600.2056-001
; CURRENT APPLICATION NUMBER: US/09/770,171
; CURRENT FILING DATE: 2001-01-29
; PRIOR APPLICATION NUMBER: US/60/178,875
; PRIOR FILING DATE: 2000-01-28
; NUMBER OF SEQ ID NOS: 1568
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1164
; LENGTH: 2124
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-770-171-1164

Alignment Scores:
Pred. No.:
Score:                  1.77e-119
Percent Similarity:     1526.00
Best Local Similarity:  99.30%
Query Match:            98.60%
DB:                     98.71%
                        33
```

```
Length:        2124
Matches:       282
Conservative:  2
Mismatches:    2
Indels:        0
Gaps:          0
```

```
US-10-054-313-1 (1-286) x US-09-770-171-1164 (1-2124)

Qy    1  MetSerTrpLeuLeuPheLeuLeuAlaHisArgValAlaLeuAlaAlaLeuAlaAlaLeuProCysArgArg   20
Db  112  ATGAGCTGGCTTCTGTTCCTGCTCCTGGCCCACAGAGTCGCCTTGGCCGCCCTGCCCTGCCGCCGC  171

Qy   21  GlySerArgGlyPheGlyMetPheTyrAlaValArgArgGlyArgLysThrGlyValPhe   40
Db  172  GGCTCTCGCGGGTTCGGGATGTTCTATGCCGTGAGGAGGGGCCGCAAGACCGGGGTCTTT  231

Qy   41  LeuThrTrpAsnGluCysArgAlaGlnValAlaAspArgArgPheProAlaAlaArgPheLysLys   60
Db  232  CTGACCTGGAATGAGTGCAGAGCACAGGTGGACCGTTTCCTGCTGCCAGATTTAAGAAG  291

Qy   61  PheAlaThrGluAspArgAlaTrpAlaPheValArgLysSerAlaSerProGluValSer   80
Db  292  TTTGCCACAGAGGATGAGGCCTGGGCCTTTGTCAGGAAAATCTGCAAGCCCGGAAGTTCA  351

Qy   81  GluGlyHisGluAsnGlnHisGlyGlnGluSerGluLeuGluProGlyLysLysArgLeuArg  100
Db  352  GAAGGGCATGAAAATCAACATGACAAGGCCGAGGCCTAGAAGCCGGAAGCGGACTCCGT  411

Qy  101  GluProLeuAspGlyAspGlyHisGluSerAlaGlnIleProTyrAlaLysHisMetLysPro  120
Db  412  GAGCCACTGGATGGAGATGGACATGAGAAAGCGCAGACCTGCAAAGCCATGAAGCCG  471

Qy  121  SerValGluProAlaProProValSerArgAspAspArgAspThrPheSerTyrMetGlyAspPheVal  140
Db  472  AGCGTGGAGCCGCCGCCCGTCTCCAGTTACAGAGACACCGTTTCCTACATGGGAGACTTCGTC  531

Qy  141  ValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgArgGlyArgLysProArgAlaGlyIle  160
Db       ...
```

Query Match: 98.71%   Indels: 0
DB: 33   Gaps: 0

US-10-054-313-1 (1-286) x US-09-785-276A-30281 (1-2129)

```
Qy     1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaAlaLeuProCysArgArg  20
Db   112 ATGAGCTGGCTTCTGTTCCTGGCCCACAGAGTCGCCTTGGCCGCCTTGCCCTGCCGCCGC 171

Qy    21 GlySerArgGlyPheGlyMetPheTyrAlaValArgArgGlyArgLysThrGlyValPhe  40
Db   172 GGCTCTCGCGGGTTCGGAATGTTCTATGCCGTGAGGAGGGGCCGAAGACCGGGGTCTTT 231

Qy    41 LeuThrTrpAsnGluCysArgAlaGlnValAlaAspArgPheProAlaAlaArgPheLysLys  60
Db   232 CTGACCTGGAATGAGTGCAGAGCACAGGTGGCAGACCGGCTTCCTGCCAGATTTAAGAAG 291

Qy    61 PheAlaThrGluAspGluAlaThrpAlaPheValArgLysSerAlaSerProGluValSer  80
Db   292 TTTGCCACAGAGGATGAGGGCCTGGCCTTTGTCAGGAAATCTGCAAGCCCGGAAGTTTCA 351

Qy    81 GluGlyHisGluAsnGlnHisGlyGlnGlnSerGluAlaLysProGlyLysArgLeuArg 100
Db   352 GAAGGGCATGAAAATCAACATGGACAAGAATCGGGAAGCCAAGCCAGGCGACTCCGT 411

Qy   101 GluProLeuAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120
Db   412 GAGCCACTGGATGGAGTGGACATGAAAGCGCAGAGCCGTATGCAAAGCACATGAAGCCG 471

Qy   121 SerValGluProAlaProProValSerArgArgAspThrPheSerTyrMetGlyAsnPheVal 140
Db   472 AGCGTGGAGCCCGGCCGCTCCAGTAGCAGAGACACGTTTCTACATGGGAGACTTCGTC 531

Qy   141 ValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGlyIle 160
Db   532 GTCGTCTACACTGATGGCTGCTGCTCCAGTAATGGGCGTAGAGGCCGGAGCAGGAATC 591

Qy   161 GlyValTyrTrpGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGln 180
Db   592 GGCGTTTACTGGGGCCAGGCCATCTTTAAATGAGGCATTAGACTTCCTGGGCGCAG 651

Qy   181 ThrAsnGlnArgAlaGluIleHisAlaAlaAlaCysLysAlaIleGluGlnAlaLysThrGln 200
Db   652 ACAAACCAAAGAGCGGAAATTCATGCAGCCTGCAAAGCCATTGAACAGCAAAGACTCAA 711

Qy   201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220
Db   712 AACATCAATAAACTGGTTCTGTATACAGACAGTATGTTTACGATAAATGGTATAACTAAC 771

Qy   221 TrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyValGluValIleAsn 240
Db   772 TGGGTTCAAGGTTGGAAGAAAAATGGTGGAAGACAAGTGCAGGGAAGAGTGATCAAC 831

Qy   241 LysGluAspPheValAlaLeuAlaLeuGluGluArgLeuThrGlnGlyMetAspIleGlnTrpHisMetHis 260
Db   832 AAAGAGGACTTTGTGGCACTGGCACTGGAAGAGAGGCTTACCCAGGGGATGGACATTCAGTGGATGCAT 891

Qy   261 ValProGlyHisSerGlyPheIleGlyValAsnGluGluAlaAspArgLeuAlaArgGluGly 280
Db   892 GTTCCTGGTCATTCGGGATTATAGGCAATGAAGAAGCTGACAGATTAGCCAGAAGGAGGA 951

Qy   281 AlaLysGlnSerGluAsp 286
Db   952 GCTAAACAATCGGAAGAC 969
```

RESULT 7
US-09-785-276A-30281
; Sequence 30281, Application US/09785276A
; GENERAL INFORMATION:
; APPLICANT: Schlegel. Robert
; APPLICANT: Endege, Wilson
; APPLICANT: Monahan, John
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF
; TITLE OF INVENTION: HUMAN PROSTATE CANCER
; FILE REFERENCE: MRI-107B
; CURRENT APPLICATION NUMBER: US/09/785,276A
; CURRENT FILING DATE: 2001-02-16
; PRIOR APPLICATION NUMBER: 60/183,319
; PRIOR FILING DATE: 2001-02-17
; PRIOR APPLICATION NUMBER: 60/189,862
; PRIOR FILING DATE: 2000-03-16
; PRIOR APPLICATION NUMBER: 60/207,454
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: 60/211,314
; PRIOR FILING DATE: 2000-06-09
; PRIOR APPLICATION NUMBER: 60/219,007
; PRIOR FILING DATE: 2000-07-18
; PRIOR APPLICATION NUMBER: 60/255,281
; PRIOR FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 62233
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 30281
; LENGTH: 2129
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2125, 2126, 2127, 2128, 2129
; OTHER INFORMATION: n = A,T,C or G
US-09-785-276A-30281

Alignment Scores:
Pred. No.:   2129
Score: 1.77e-119
Percent Similarity: 1526.00   Matches: 282
Best Local Similarity: 99.30%   Conservative: 2
                         98.60%   Mismatches: 2

RESULT 8
US-10-357-930-30281
; Sequence 30281, Application US/10357930
; GENERAL INFORMATION:
; APPLICANT: Schlegel. Robert
; APPLICANT: Endege, Wilson
; APPLICANT: Monahan, John
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR

; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF
; TITLE OF INVENTION: HUMAN PROSTATE CANCER
; FILE REFERENCE: MRI-007BCN
; CURRENT APPLICATION NUMBER: US/10/357,930
; CURRENT FILING DATE: 2003-02-04
; PRIOR APPLICATION NUMBER: 09/785,276
; PRIOR FILING DATE: 2003-02-16
; PRIOR APPLICATION NUMBER: 60/183,319
; PRIOR FILING DATE: 2000-02-17
; PRIOR APPLICATION NUMBER: 60/189,862
; PRIOR FILING DATE: 2000-03-16
; PRIOR APPLICATION NUMBER: 60/207,454
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: 60/211,314
; PRIOR FILING DATE: 2000-06-09
; PRIOR APPLICATION NUMBER: 60/219,007
; PRIOR FILING DATE: 2000-07-18
; PRIOR APPLICATION NUMBER: 60/255,281
; PRIOR FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 6232
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 30281
; LENGTH: 2129
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2125, 2126, 2127, 2128, 2129
; OTHER INFORMATION: n = A,T,C or G
US-10-357-930-30281

Alignment Scores:
Pred. No.:   Length: 2129
Score: 1.77e-119   Matches: 282
                        1526.00
Percent Similarity: 99.30%   Conservative: 2
Best Local Similarity: 98.60%   Mismatches: 2
Query Match: 98.71%   Indels: 0
DB: 51   Gaps: 0

US-10-054-313-1 (1-286) x US-10-357-930-30281 (1-2129)

Qy   1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaLeuAlaAlaLeuProCysArgArg 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 112 ATGAGCTGGCTTCGTTCTTCCTGGCCCACAGAGTCGCCCTGGCCCTTGCCCTGCCCTGCCCGCCGC 171

Qy   21 GlySerArgGlyPheGlyMetPheTyrAlaValArgArgGlyArgLysThrGlyValPhe 40
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db 172 GGCTCTCGCGGGTTCGGGATGTTCTATGCCGTGAGGAGGGGCCGCAAGACCGGGGTCTTT 231

Qy   41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys 60
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db 232 CTGACCTGGAATGAGTGCAGAGCACAGGTGGACCGGTTTCCTGCTGCCAGATTTAAGAAG 291

Qy   61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db 292 TTTGCCACAGAGGATGAGGCCTGGGCCTTTGTCAGGAAATCTGCAAGCCCGGAAGTTTCA 351

Qy   81 GlyGlyHisGluAsnGlnHisGlyGlyAsnGlnSerGlyAlaLysProGlyLysArgLeuArg 100
          ||||||||||||||||||||||||||||||||||||||
Db 352 GAAGGGCATGAAAATCAACATGGCAAGAATCGGAGGGGAAAGCCAGCAAGCGACTCCGT 411

Qy 101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120
          |||||||::|||||||||||||||||||||
Db 412 GAGCCACTGGATGGAGATGGACATGAAAGCGCAGCCGTATGCAAAGCACATGAAGCCG 471

Qy 121 SerValGluProAlaProProValSerArgArgAspThrPheSerSerTyrMetGlyAspPheVal 140
          ||||||||||||||||||||||||||||||||||
Db 472 AGCGTGGAGCCGGCGCCGCCTGTCTCCAGGAGAGACACGTTTTCCTACATGGGAGACTTCGTC 531

Qy 141 ValValTyrThrAspArgCysCysSerSerAsnGlyArgArgLysProArgAlaGlyIle 160
          ||||||||||||::|||||||||||:::||
Db 532 GTCGTCTACACTGATGGCTGCTGCTCCAGTAATGGGCGTAGAAGGCCCGCAGCAGGAATC 591

Qy 161 GlyValTyrTrpGlyTyrProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGln 180
          |||||||||||||||||||||||||||||||||
Db 592 GGCGTTTACTGGGGCTACCCAGGCCATCCTTTAAATGTGGGCATTAGACTTCCTGGGCGCAG 651

Qy 181 ThrAsnGlnArgAlaGluHisAlaAlaCysLysAlaLysThrGln 200
          ||||||||||||||||||||||||||
Db 652 ACAAACCAAAGAGCGGGAAATTCATGCAGCCTCCAAAGCCATTGAACAAGCAAAGACTCAA 711

Qy 201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220
          ||||||||||||||||||||||||||||||||||
Db 712 AACATCAATAAACTGGTTCTGTATACAGACAGTATGTTTACGATAAATGGTATAAACTAAC 771

Qy 221 TrpValGlnGlyTrpLysAsnGlyTrpLysThrSerAlaGlyValLysGluValIleAsn 240
          ||||||||||||||||||||||||||||||||||
Db 772 TGGGTTCAAGGTTGGAAGAAATGGGTGGAAGACAAGTGCAGGGGGAAAGAGTGATCAAC 831

Qy 241 LysGluAspPheValAlaLeuGluArgLeuThrGlyMetGlyAspIleGlnTrpMetHis 260
          ||||||||||||||||||||||||||||||||||
Db 832 AAAGAGGACTTTGTGCCACTGGAGAGCTTACCCAGGGATGGACATTCAGTGGATGCAT 891

Qy 261 ValProGlyHisSerGlyPheIleGlyAsnGluGluAlaArgLeuAlaArgGluGluGly 280
          ||||||||||||||||||||||||||||||||||
Db 892 GTTCCTGGTCATTCGGGATTTATAGGCAATGAAGAAGAGCTGACAGATTAGCCAGAGGA 951

Qy 281 AlaLysGlnSerGluAsp 286
          ||||||||||||
Db 952 GCTAAACAATCGGAAGAC 969

RESULT 9
US-09-763-233-41
; Sequence 41, Application US/09763233
; GENERAL INFORMATION:
; APPLICANT: INCYTE PHARMACEUTICALS, INC.
; APPLICANT: HILLMAN, Jennifer L.
; APPLICANT: YUE, Henry
; APPLICANT: TANG, Y. Tom
; APPLICANT: CORLEY, Neil C.
; APPLICANT: GUEGLER, Karl J.
; APPLICANT: GORGONE, Gina A.
; APPLICANT: PATTERSON, Chandra
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: LAL, Preeti
; APPLICANT: BANDMAN, Olga
; APPLICANT: REDDY, Roopa
; APPLICANT: AZIMZAI, Yalda
; APPLICANT: SHIH, Leo L.
; APPLICANT: YANG, Junming
; APPLICANT: LU, Dyung Aina M.
; TITLE OF INVENTION: HUMAN RNA-ASSOCIATED PROTEINS
; FILE REFERENCE: PF-0579 PCT
; CURRENT APPLICATION NUMBER: US/09/763,233
; CURRENT FILING DATE: 2002-08-26
; PRIOR APPLICATION NUMBER: 60/097,550; 60/115,639
; PRIOR FILING DATE: 1998-08-21; 1999-01-12
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PERL Program
; SEQ ID NO 41
; LENGTH: 1150
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte Identification No.: 2073417CB1
US-09-763-233-41

Alignment Scores:
Pred. No.:   Length: 1150
Score: 1.59e-119   Matches: 281
                        1523.00
Percent Similarity: 99.30%   Conservative: 3
Best Local Similarity: 98.25%   Mismatches: 2
Query Match: 98.51%   Indels: 0
DB: 33   Gaps: 0

APPLICANT: Yalda Azimzai
APPLICANT: Leo Shih
APPLICANT: Junming Yang
APPLICANT: Aina Lu
TITLE OF INVENTION: HUMAN RNA-ASSOCIATED PROTEINS
FILE REFERENCE: PF-0655 P
CURRENT APPLICATION NUMBER: US/60/115,639
CURRENT FILING DATE: 1999-01-12
NUMBER OF SEQ ID NOS: 42
SOFTWARE: PERL Program
SEQ ID NO 32
LENGTH: 1150
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE: -
OTHER INFORMATION: 2073417
US-60-115-639-32

Alignment Scores:
Pred. No.:
Score: 1.59e-119
Percent Similarity: 1523.00
Best Local Similarity: 99.30%
Query Match: 98.25%
DB: 98.51%
                               65

Length: 1150
Matches: 281
Conservative: 3
Mismatches: 2
Indels: 0
Gaps: 0

US-60-115-639-32

US-10-054-313-1 (1-286) x US-60-115-639-32 (1-1150)

US-10-054-313-1 (1-286) x US-09-763-233-41 (1-1150)

RESULT 10
US-60-115-639-32
; Sequence 32, Application US/60115639
; GENERAL INFORMATION:
; APPLICANT: Preeti Lal
; APPLICANT: Henry Yue
; APPLICANT: Olga Bandman
; APPLICANT: Y. Tom Tang
; APPLICANT: Jennifer L. Hillman
; APPLICANT: Roopa Reddy
; APPLICANT: Mariah R. Baughn

```
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 439
;  SOFTWARE: CuraSeqList version 0.1
;  SEQ ID NO 85
;    LENGTH: 965
;    TYPE: DNA
;    ORGANISM: CG144997-02
;  FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (10)..(867)
PCT-US02-31373-85

Alignment Scores:
Pred. No.:
Score:                1.58e-119      Length:         965
                      1522.00        Matches:        281
Percent Similarity:   98.95%         Conservative:   2
Best Local Similarity: 98.25%        Mismatches:     3
Query Match:          98.45%         Indels:         0
DB:                   1              Gaps:           0

US-10-054-313-1 (1-286) x PCT-US02-31373-85 (1-965)

Qy    1 MetSerTrpLeuPheLeuAlaHisArgValAlaLeuAlaAlaLeuProCysArgArg  20
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10 ATGAGCTGGTTGTTCTGTTCCTGGCCCACAGAGTCGCCTTGGCCGCCTTGCCGCCGC  69

Qy   21 GlySerArgGlyPheMetPheTyrAlaValArgGlyArgGlyLysThrGlyValPhe  40
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   70 GGCTCTCGCGGGTTCATGTTCTGGGATGTTCTATGCCGTGAGAGGGCGCAAGACCGGGGTCTTT 129

Qy   41 LeuThrTrpAsnGluCysArgAlaGlnValAspAsnArgPheProAlaAlaArgPheLysLys  60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  130 CTGACCTGGAATGAGTGCAGAGCACAGGTGGACCGGTTTCCTGCTGCCAGATTTAAGAAG  189

Qy   61 PheAlaThrGluAspGluAlaTrpAlaTrpLysSerProGluValSer  80
          |||||||||||||||||||||||||||||||||||||||||||||||
Db  190 TTTGCCACAGAGGATGAGGCCTGGGCCTTTGTCAGGAAATCTGCAAGCCCGGAAGTTTCA  249

Qy   81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg  100
          |||||||||||||||||||||||||||||||||||||||||||||||
Db  250 GAAGGGCATGAAAATCAACATGGACAAGAATCGGAGGCAACAAGCAAGCAAGCGACTCCGT  309

Qy  101 GluProLeuAspGlyHisGluAspGlnProTyrAlaLysHisMetLysPro  120
          ||||||||||||||||||::|||||||||||||||||||||||||
Db  310 GAGCCACTGGATGGACATGGAAGACCAGCCGTATGCAAAGCATATGAAGCCG  369

Qy  121 SerValGluProAlaProProValSerArgAspThrPheSerTyrMetGlyAspPheVal  140
          |||||||||||||||||||||||||||||||||||||||||||||||
Db  370 AGCGTGGAGCCGGCGCCTCCAGTTAGCAGAGACACAGTTTCTTACATGGGAGACTTCGTC  429

Qy  141 ValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGlyIle  160
          |||||||||||||||||||||||||||||||:::||||||||||||
Db  430 GTCGTCTACACTGATGGCTGCTGCTCCAGTAATGGCCGTAGAAGGCCGCGAGCAGGAATC  489

Qy  161 GlyValTyrTyrProGlyHisProLeuAsnValGlyTyrAlaArgLeuProGlyArgGln  180
          |||||||||||||||||||||||||||||||||||||||||||||||
Db  490 GGCGTTTACTGGGGCCAGCCATCTTAAATGTGGCATTAGGGATCTTGGGGCGGCAG  549

Qy  181 ThrAsnGlnArgGluAlaIleHisAlaAlaCysLysAlaIleGluGlnAlaThrGln  200
          |||||||||||||||||||||||||||||||||||||||||||||||
Db  550 ACAAACCAAAGAGGGCGAAATTCATGCAGCCTGCAAAGCCAGCCATTGAACAAGCAAAGACTCAA 609

Qy  201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn  220
          |||||||||||||||||||||||||||||||||||||||||||||||
Db  610 AACATCAATAAACTGGTTCTGTATACAGACAGTATGTTTACGATAAATGGTATAACTAAC  669

Qy  221 TrpValTyrTrpLysTrpLysAsnGlyTrpLysSerThrSerAlaGlyLysValIleAsn  240
          |||||||||||||||||||||||||||||||||||||||||||||||
Db  670 TGGGTTCAAGGTTGGAAGAAAATGGGTGGAAGAAGTGCAGGGAAAGAGGTGATCAAC  729

Qy  241 LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHis  260
          |||||||||||||||||||||||||||||||||||||||||||||||
Db  730 AAAGAGGACTTTGTGCCACTGGAGAGCTTACCCAGGGGATGGACATTCAGTGGATGCAT  789

          |||||||||||||||||||||||||||||||||||||||||||||||
Db  745 TGGGTTCAAGGTTGGAGAAAATGGGTGGAGACAAGTCGCAGGGAAAGAGAGTGATCAAC  804

Qy  241 LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHis  260
          |||||||||||||||||||||||||||||||||||||||||||||||
Db  805 AAAGAGGACTTTGTGCCACTGGAGAGCTTACCCAGGGGATGGACATTCAGTGGATGCAT  864

Qy  261 ValProGlyHisSerGlyPheIleGlyLysAsnGluGluAlaAspArgLeuAlaArgGluGlyGly 280
          |||||||||||||||||||||||||||||||||||||||||||||||
Db  865 GTTCCTGGTCAATCGGGATTATAGGCAATGAAGAAGAGAGCTGACAGATTAGCCAGGAAGGA  924

Qy  281 AlaLysGlnSerGluAsp  286
          ||||||||||||||||||
Db  925 GCTAAACAATCGGAAGAC  942
```

RESULT 11
PCT-US02-31373-85

```
Qy   261  ValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgGluGly  280
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   790  GTTCCTGGTCATTCGGGATTTATAGGCAATGAACAATGAAGAAGCTGACAGATTAGCCAGAGAAGGA  849

Qy   281  AlaLysGlnSerGluAsp  286
          ||||||||||||||||||
Db   850  GCTAAACAATCGGAAGAC  867
```

RESULT 12
US-10-262-511-85
; Sequence 85, Application US/10262511
GENERAL INFORMATION:

APPLICANT: Smithson, Glenda
APPLICANT: Millet, Isabelle
APPLICANT: Peyman, John A.
APPLICANT: Kekuda, Ramesh
APPLICANT: Ju, Jingfang
APPLICANT: Li, Li
APPLICANT: Guo, Xiaojia (Sasha)
APPLICANT: Pasturajan, Meera
APPLICANT: Spytek, Kimberly A.
APPLICANT: Edinger, Shlomit R.
APPLICANT: Ellerman, Karen
APPLICANT: Malyankar, Uriel M.
APPLICANT: Ort, Tatiana
APPLICANT: Gorman, Linda
APPLICANT: Zerhusen, Bryan D.
APPLICANT: Anderson, David W.
APPLICANT: Zhong, Mei
APPLICANT: Catterton, Elina
APPLICANT: Ji, Weizhen
APPLICANT: Miller, Charles E.
APPLICANT: Rastelli, Luca
APPLICANT: Stone, David J.
APPLICANT: Pena, Carol E. A.
APPLICANT: Shenoy, Suresh G.
APPLICANT: Shimkets, Richard A.
APPLICANT: Rothenberg, Mark E.
APPLICANT: Leach, Martin D.
APPLICANT: Agee, Michele L.
APPLICANT: Berghs, Constance
TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
FILE REFERENCE: 21402-462C
CURRENT APPLICATION NUMBER: US/10/262,511
CURRENT FILING DATE: 2003-05-28
PRIOR APPLICATION NUMBER: 60/326,483
PRIOR FILING DATE: 2001-10-02
PRIOR APPLICATION NUMBER: 60/373,815
PRIOR FILING DATE: 2002-04-19
PRIOR APPLICATION NUMBER: 60/327,917
PRIOR FILING DATE: 2001-10-09
PRIOR APPLICATION NUMBER: 60/381,642
PRIOR FILING DATE: 2002-05-17
PRIOR APPLICATION NUMBER: 60/328,029
PRIOR FILING DATE: 2002-10-09
PRIOR APPLICATION NUMBER: 60/381,038
PRIOR FILING DATE: 2002-05-16
PRIOR APPLICATION NUMBER: 60/328,056
PRIOR FILING DATE: 2001-10-09
PRIOR APPLICATION NUMBER: 60/373,260
PRIOR FILING DATE: 2002-04-17
PRIOR APPLICATION NUMBER: 60/373,826
PRIOR FILING DATE: 2002-04-19
PRIOR APPLICATION NUMBER: 60/327,435
PRIOR FILING DATE: 2001-10-05
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 439
SOFTWARE: CuraSeqList version 0.1
SEQ ID NO 85
LENGTH: 965
TYPE: DNA
ORGANISM: Homo sapiens

; FEATURE:
; NAME/KEY: CDS
; LOCATION: (10)..(867)
US-10-262-511-85

Alignment Scores:
Pred. No.:   1.58e-119
Score: 1522.00
Percent Similarity: 98.95%
Best Local Similarity: 98.25%
Query Match: 98.45%
DB: 49

Length: 965
Matches: 281
Conservative: 2
Mismatches: 3
Indels: 0
Gaps: 0

```
US-10-054-313-1 (1-286) x US-10-262-511-85 (1-965)

Qy     1  MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaAlaAlaLeuProCysArgArg   20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    10  ATGAGCTGGTTTCTGTTCCTGGCCCACAGAGTCGCCTTGGCCGCCGCCCTGCCCTGCCGCCGC   69

Qy    21  GlySerArgGlyMetPheTyrAlaValArgArgGlyArgGlyThrGlyValPhe   40
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    70  GGCTCTCGCGGGTTCGGGATGTTCTATGCCGTGAGGAGGGGCCGAAGACCGGGGTCTTT  129

Qy    41  LeuThrTrpAsnGluCysArgArgAlaGlnIleValAlaSpArgPheProAlaAlaArgPheLysLys   60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   130  CTGACCTGGAATGAGTGCAGAGCACAGTGGACCGGTTCCTGCTGCCAGATTTAAGAAG  189

Qy    61  PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer   80
          |||||||||||||||||||||||||||||||||||||||||||||||
Db   190  TTTGCCACAGACGATGAGGCCCTGGCCTTTGTCAGGAAATCTGCAAGCCCGGAAGTTTCA  249

Qy    81  GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg  100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   250  GAAGGGCATGAAAATCAACATGGACAAGAAATCGGAGGCAGCCAGCAAGCGACTCCGT  309

Qy   101  GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro  120
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   310  GAGCCACTGGATGGAGATGGACATGAAAGCGCAGAGCCGTATGCAAAGCACATGAAGCCG  369

Qy   121  SerValGluProAlaProValSerArgAspThrPheSerTyrMetGlyAspPheVal  140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   370  AGCGTGGAGCCCGCGCCTCCAGTTAGCAGAGACACGTTTTCTACATGGGAGACTTCGTC  429

Qy   141  ValValTyrThrAspIleCysCysSerSerArgAsnGlyArgArgProArgAlaGlyIle  160
          |||||||||||||||||||||||::|:|||||||||||||||::|||||:::||||
Db   430  GTCGTCTACACTGATGCGTGCTGTCTCCAGTAATGGGCGTAGAAGGCCGCGAGCCAGGAATC  489

Qy   161  GlyValTyrTrpProGlyHisProLeuAsnValGlyTyrIleArgLeuProGlyArgGln  180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   490  GGCGTTTACTGGGGGCCAGGCCATCCTTAAATGTAGGCATTAGACTTCCTGGCGGCAG  549

Qy   181  ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaIleGluGlnAlaLysThrGln  200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   550  ACAAACCAAAGAGCGGAAATTCATGCAGCCTGCAAAGCCATTGAACAAGCAAAGACTCAA  609

Qy   201  AsnIleAsnLysGluValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn  220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   610  AACATCAATAAACTGGTTCTGTATACAGACAGTATGTTTACGATAAATGTATAAACTAAC  669

Qy   221  TrpValGlnGlyTrpLysTyrAsnGlyTyrTrpLysThrSerAlaLeuGlyValIleAsn  240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   670  TGGGTTCAAGGTTGGAAGAAAAATGGGTGGAAGACAAGTGCAGGGAAAGAGGTGATCAAC  729

Qy   241  LysGluAspPheValAlaLeuGluThrGlnGlyMetAspIleGlnTrpMetHis  260
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   730  AAAGAGGACTTTGTGGCACTGGAGACGGCTTACCAGGGGGATGGACATTCAGTGGATGCAT  789

Qy   261  ValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgGluGly  280
          |||||||||||||||||||||||||||||||||||||||||
Db   790  GTTCCTGGTCATTCGGGATTTATAGGCAATGAACAATGAAGAAGCTGACAGATTAGCCAGAGAAGGA  849

Qy   281  AlaLysGlnSerGluAsp  286
          ||||||||||||
```

```
Db        850 GCTAAACAATCGGAAGAC 867

RESULT 13
US-10-262-511A-85
; Sequence 85, Application US/10262511A
; GENERAL INFORMATION:
;   APPLICANT: Smithson, Glennda
;   APPLICANT: Millet, Isabelle
;   APPLICANT: Peyman, John A.
;   APPLICANT: Kekuda, Ramesh
;   APPLICANT: Ju, Jingfang
;   APPLICANT: Li, Li
;   APPLICANT: Guo, Xiaojia (Sasha)
;   APPLICANT: Patturajan, Meera
;   APPLICANT: Spytek, Kimberly A.
;   APPLICANT: Edinger, Shlomit R.
;   APPLICANT: Ellerman, Karen
;   APPLICANT: Malyankar, Uriel M.
;   APPLICANT: Ort, Tatiana
;   APPLICANT: Gorman, Linda
;   APPLICANT: Zerhusen, Bryan D.
;   APPLICANT: Anderson, David W.
;   APPLICANT: Zhong, Mei
;   APPLICANT: Catterton, Elina
;   APPLICANT: Ji, Weizhen
;   APPLICANT: Miller, Charles E.
;   APPLICANT: Rastelli, Luca
;   APPLICANT: Stone, David J.
;   APPLICANT: Pena, Carol E. A.
;   APPLICANT: Shenoy, Suresh G.
;   APPLICANT: Shimkets, Richard A.
;   APPLICANT: Rothenberg, Mark E.
;   APPLICANT: Leach, Martin D.
;   APPLICANT: Agee, Michele L.
;   APPLICANT: Berghs, Constance
;   TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
;   FILE REFERENCE: 21402-462C
;   CURRENT APPLICATION NUMBER: US/10/262,511A
;   CURRENT FILING DATE:  2002-10-01
;   PRIOR APPLICATION NUMBER: 60/326,483
;   PRIOR FILING DATE:  2001-10-02
;   PRIOR APPLICATION NUMBER: 60/373,815
;   PRIOR FILING DATE:  2002-04-19
;   PRIOR APPLICATION NUMBER: 60/327,917
;   PRIOR FILING DATE:  2001-10-09
;   PRIOR APPLICATION NUMBER: 60/381,642
;   PRIOR FILING DATE:  2002-05-17
;   PRIOR APPLICATION NUMBER: 60/328,029
;   PRIOR FILING DATE:  2002-10-09
;   PRIOR APPLICATION NUMBER: 60/381,038
;   PRIOR FILING DATE:  2002-05-16
;   PRIOR APPLICATION NUMBER: 60/328,056
;   PRIOR FILING DATE:  2001-10-09
;   PRIOR APPLICATION NUMBER: 60/373,260
;   PRIOR FILING DATE:  2002-04-17
;   PRIOR APPLICATION NUMBER: 60/373,826
;   PRIOR FILING DATE:  2002-04-19
;   PRIOR APPLICATION NUMBER: 60/327,435
;   PRIOR FILING DATE:  2001-10-05
;   Remaining Prior Application data removed - See File Wrapper or PALM.
;   NUMBER OF SEQ ID NOS: 439
;   SOFTWARE: CuraSeqList version 0.1
SEQ ID NO 85
;   LENGTH: 965
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (10)..(867)
US-10-262-511A-85

Alignment Scores:
Pred. No.:                                        1.58e-119       Length:       965
```

```
Score:                      1522.00
Percent Similarity:         98.95%        Matches:       281
Best Local Similarity:      98.25%        Conservative:    2
Query Match:                98.45%        Mismatches:      3
DB:                                       Indels:          0
                                          Gaps:            0

US-10-054-313-1 (1-286) x US-10-262-511A-85  (1-965)

Qy    1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaAlaLeuProCysArgArg  20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10 ATGAGCTGGTTGTTCCTGGCCCACAGAGTCGCCCTGGCCGCCCTGCCCTGCCGCGCCGC  69

Qy   21 GlySerArgGlyPheGlyMetPheTyrAlaValArgGlyLysThrGlyValValPhe  40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   70 GGCTCTCGCGGGTTCGGGATGTTCTATGCCGTGAGGGGCCCGAAGACCGGGGTCTTT  129

Qy   41 LeuThrTrpAsnLeuCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys  60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  130 CTGACCTGGAATGAGTGCAGAGCACAGGTGGACCGGTTTCCTGCTGCCAGATTTAAGAAG  189

Qy   61 PheAlaThrGluAspGluAlaTrpAlaArgLysSerAlaSerProGluValSer  80
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  190 TTTGCCACAGAGGATGAGGCTGGGCCTTTGTCAGGAAATCTGGAAGTTTCA  249

Qy   81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysProGlyLeuArg  100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  250 GAAGGGCATGAAAATCAACATGGACAAGAATCGGAGGCAAGAGCCAGCAAGCAGACTCCGT  309

Qy  101 GluProLeuAspGlyHisAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro  120
          ||||||||||||||||||||||||||| |||||| |||||||||||||||||||||||
Db  310 GAGCCACTGGATGGACATGAAGGCCAGCCATGTCAAAGCACATGAAAGCCG  369

Qy  121 SerValGluProAlaProProValSerArgAspThrPheSerTyrMetGlyAspPheVal  140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  370 AGCGTGGAGCCGGCGCCTCCAGTAGCAGACACGTTTTCCTACATGGGAGACTTCGTC  429

Qy  141 ValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGlyIle  160
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  430 GTCGTCTACACTGATGGCTGCTGCTCCAGTAATGGCCGTAGAGGCCCGAGCCAGCAGAATC  489

Qy  161 GlyValTyrTrpGlyTrpProLeuAsnValGlyLeuProGlyValArgGln  180
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  490 GGGGTTTACTGGGGCTGGCCACTCCTTTAAATGTAGGCATTAGGCCGGGCCAG  549

Qy  181 ThrAsnGlnArgAlaGluIleHisAlaAlaCysGlyAlaIleGluGluGlnAlaLysThrGln  200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  550 ACAAACCAAAGAGCGGCGGAAATTCATGCAGCCTGCAAAGCCATTGACAAGCAAAGACTCAA  609

Qy  201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleThrIleThrAsn  220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  610 AACATCAATAAACTGGTTCTGTATACAGACAGTATGTTTACGATAAATGGTATAAACTAAC  669

Qy  221 TrpValGlnGlyTrpProLysLysAsnGlyTrpLysThrSerAlaGlyLysGluValIleAsn  240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  670 TGGGTTCAAGGTTGGAAGAAAAATGGGTGGAAGACAAGTGCAGGGAAAGAGGTGATCAAC  729

Qy  241 LysGluAspPheValAlaLeuGluThrGlnTrpMetAspIleGlnTrpMetHis  260
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  730 AAAGAGGACTTTGTGCCACTGGAGACGCTTACCCAGGGATGGATGCATTCAGTGGATGCAT  789

Qy  261 ValProGlyHisSerGlyPheIleGlyPheAsnGluGluGluAlaAspArgLeuAlaArgGluGlyGly  280
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  790 GTTCCTGGTCATTCGGGATTTATAGGCAATGAACAGATTAGCCAGAGAAGGA  849

Qy  281 AlaLysGlnSerGluAsp  286
          ||||||
Db  850 GCTAAACAATCGGAAGAC  867

RESULT 14
PCT-US02-18947-841
; Sequence 841, Application PC/TUS0218947
; GENERAL INFORMATION:
;   APPLICANT: Rosetta Inpharmatics
```

; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-175-228
; CURRENT APPLICATION NUMBER: PCT/US02/18947
; CURRENT FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/380,770
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 841
; LENGTH: 1147
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: NM_002936
; DATABASE ENTRY DATE: 2001-06-18
PCT-US02-18947-841

Alignment Scores:
Pred. No.:   Length: 1147
Score: 1522.00   Matches: 281
Percent Similarity: 98.95%   Conservative: 2
Best Local Similarity: 98.25%   Mismatches: 3
Query Match: 98.45%   Indels: 0
DB: 1   Gaps: 0

US-10-054-313-1 (1-286) x PCT-US02-18947-841 (1-1147)

Qy   1   MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaAlaLeuProCysArgArg   20
Db   62   ATGAGCTGGTTGTTCCTGGCCCACAGAGTCGCCTTGGCCGCCCTTGCCCTGCCGCGCCGC   121

Qy   21   GlySerArgGlyPheGlyMetPheTyrAlaValArgArgGlyArgLysThrGlyValValPhe   40
Db 122   GGCTCTCGCGGGTTCGGGATGTTCTATGCCGTGAGGAGGGGCCGCAAGACCGGGGTCTTT   181

Qy   41   LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys   60
Db 182   CTGACCTGGAATGAGTGCAGAGCACAGGTGGACCGGTTTCCTGCTGCCAGATTTAAGAAG   241

Qy   61   PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer   80
Db 242   TTTGCCACAGAGGATGAGCCTTGGCCCTTTGTCAGAAAATCTGCAAGCCCGGAAGTTTCA   301

Qy   81   GluGluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg   100
Db 302   GAAGGGCATGAAAATCAACATGGACAAGAATCGGAGGGAAAGCCAGCAAGCGACTCCGT   361

Qy 101   GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro   120
Db 362   GAGCCACTGGACGGGGATGGACATGAAAAGCGCAGACCCTATGCAAAGCACATGAAGCCG   421

Qy 121   SerValGlyProLeuAlaProProValSerArgAspThrPheSerTyrMetGlyAspPheVal   140
Db 422   AGCGTGGGAGCCGGGCCCTCCAGTTAGCAGAGACAGACGTTTTTCCTACATGGGAGACTTCGTC   481

Qy 141   ValValTyrThrAspLysCysCysSerSerSerAsnGlyArgArgGlyArgLysProArgAlaGlyIle   160
Db 482   GTCGTCTACACTGATGAGTGCTGTCTCCAGTAATGGAAGGCGCGGAGCCAGGAATC   541

Qy 161   GlyValTyrThrProGlyHisProLeuAsnValGlyIleLeuArgLeuProGlyArgGln   180
Db 542   GGCGTTTACTGGGGGCCAGGCCATCCTTTAAATGTAGGCATTAGGACTTCCTGGCGCCAG   601

Qy 181   ThrAsnGlnArgAlaGluIleHisAlaCysLysLysAlaIleGluGluHisAlaLysThrGln   200
Db 602   ACAAACCAAAGAGCCGAAATTCATGCAGCCTGCAAAAGGCCATTGAACAAGCAAAGACTCA   661

Qy 201   AsnIleAsnLysAsnValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn   220
Db 662   AACATCAATAAAACTGTTCTGTATACAGACAGTATGTTTACGATAAATGTATAACTAAC   721

Qy 221   TrpValGlnGlyTyrGlyLysAsnGlyTrpLysSerThrSerAlaGlyTrpThrLysLysAspProGlnValIleAsn   240
Db 722   TGGGTTCAAGGTTGGAAGAAAAATGGTGGAAGCAAGTGCAGGGAAGAGGTGATCAAC   781

; (signature)

RESULT 15
US-10-172-118-841
; Sequence 841, Application US/10172118
; GENERAL INFORMATION:
;   APPLICANT: Dai, Hongyue
;   APPLICANT: He, Yudong
;   APPLICANT: Linsley, Peter
;   APPLICANT: Mao, Mao
;   APPLICANT: Roberts, Chris
;   APPLICANT: Van 't Veer, Laura
;   APPLICANT: Van de Vijver, Marc
;   APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-175-999
; CURRENT APPLICATION NUMBER: US/10/172,118
; CURRENT FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/380,770
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 841
; LENGTH: 1147
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: NM_002936
; DATABASE ENTRY DATE: 2001-06-18
US-10-172-118-841

Alignment Scores:
Pred. No.:   Length: 1147
Score: 1522.00   Matches: 281
Percent Similarity: 98.95%   Conservative: 2
Best Local Similarity: 98.25%   Mismatches: 3
Query Match: 98.45%   Indels: 0
DB: 47   Gaps: 0

US-10-054-313-1 (1-286) x US-10-172-118-841 (1-1147)

Qy   1   MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaAlaLeuProCysArgArg   20
Db   62   ATGAGCTGGTTGTTCCTGTTCCTGGCCCACAGAGTCGCCTTGGCCGCCCTTGCCCTGCCGCGCCGC   121

Qy   21   GlySerArgGlyMetPheTyrAlaValArgArgGlyArgLysThrGlyValValPhe   40
Db 122   GGCTCTCGCGGGTTCGGGATGTTCTATGCCGTGAGGAGGGGCCGCAAGACCGGGGTCTTT   181

Qy   41   LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys   60
Db 182   CTGACCTGGAATGAGTGCAGAGCACAGGTGGACCGGTTTCCTGCTGCCAGATTTAAGAAG   241

Qy   61   PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer   80
Db 242   TTTGCCACAGAGGATGAGGCCTGGGCCTTTGTCAGGAAAATCTCAGGCCCGGAAGTTTCA   301

Qy   81   GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg   100
Db 302   GAAGGGCATGAAAATCAACATGGACAAGAATCGGAGGGAAAGCCAGCAAGCGACTCCGT   361

Qy 101   GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro   120
Db 302   GAGGGGCATGAAAATCAACATGGACAAGAATCGGAGGGAAAGCCAGCAAGCGACTCCGT   361

Qy 241   LysGluAspPheValAlaAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHis   260
Db 782   AAAGAGGACTTTGTGGCACTGGAGAGGCTTACCCAGGGGATGGACATTCAGTGATGCAT   841

Qy 261   ValProGlyHisSerGlyPheIleGlyLysAsnGluGluAlaAspArgLeuAlaArgGluGly   280
Db 842   GTTCCTGGTCATTCGGATTTATAGGCAATGAAGAAGCTGACAGATTAGCCAGAGAAGGA   901

Qy 281   AlaLysGlnSerGluAsp   286
Db 902   GCTAAACAATCGGAAGAC   919

```
Db  362 GAGCCACTGGATGGAGATGGACATGAAAGCGCGAGAGCCGTATGCAAAGCACATGAAGCCG 421

Qy  121 SerValGluProAlaProValSerArgAspThrPheSerTyrMetGlyAspPheVal 140
        |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  422 AGCGTGGAGCCCGGCCGCCTCCAGTTAGCAGAGACACGTTTCCTACATGGGAGACTTCGTC 481

Qy  141 ValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGlyIle 160
        ||||||||||||||||||||||||||||||||:::||||||||||||||||||||
Db  482 GTCGTCTACACTGATGGCTGCTGCTCCAGTAATGGGCGTAGAAGGCCGGAGCAGGAATC 541

Qy  161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyIleArgIleArgGlyArgGln 180
        |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  542 GGCGTTTACTGGGGCCCAGGGCCATCCTTTAAATGTAGCATTAGACTTCCTGGGCGCAG 601

Qy  181 ThrAsnGlnArgAlaGluHisAlaAlaCysLysAlaIleGluGlnAlaLysThrGln 200
        ||||||||||||||||||||||||||||||||||||||||||||||||
Db  602 ACAAACCAAAGAGCGGAAATTCATGCAGCCTGCAAAGCCATTGAACAAGCAAAGACTCAA 661

Qy  201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220
        |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  662 AACATCAATAAACTGGTTCTGTATACAGACAGTATGTTTACGATAAATGGTATAACTAAC 721

Qy  221 TrpValGlnGlyTrpLysAsnGlyTrpLysTrpSerAlaGlyLysGluValIleAsn 240
        ||||||||||||||||||||||||||||||||||||||||||||||||
Db  722 TGGGTTCAAGGTGGAAGAAAAATGGGTGGAAGACAAGTGCAGGGAAAGAGGTGATCAAC 781

Qy  241 LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHis 260
        |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  782 AAAGAGGACTTTGTGGCACTGGAGAGGCTTACCCAGGGGATGGACATTCAGTGGATGCAT 841

Qy  261 ValProGlyHisSerGlyPheIleGlyAsnGluGlyAlaAlaAspArgLeuAlaArgGluGly 280
        ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  842 GTTCCTGGTCATTCGGGATTTATAGGCAATGAAGAAGGCTGACAGATTAGCCAGAGAAGGA 901

Qy  281 AlaLysGlnSerGluAsp 286
        ||||||||||||||
Db  902 GCTAAACAATCGGAAGAC 919
```

Search completed: December 17, 2003, 22:11:56
Job time : 3554 secs

OM protein - nucleic search, using frame_plus_p2n model

Copyright   GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

Run on: December 17, 2003, 18:09:14 ; Search time 2512 Seconds
                                (without alignments)
                                2767.150 Million cell updates/sec

Title: US-10-054-313-1
Perfect score: 1546
Sequence: 1 MSWLFLPAHRVALAALPCRR........FIGNEEADRLARBGAKQSED 286

Scoring table: BLOSUM62
                Xgapop 10.0 , Xgapext 0.5
                Ygapop 10.0 , Ygapext 0.5
                Fgapop 6.0 , Fgapext 7.0
                Delop   6.0 , Delext   7.0

Searched: 22781392 seqs, 1215223056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10054313/runat_17122003_150744_24191/app_query.fasta_1.455
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10054313 @CGN_1 1 2810 @runat_17122003_150744_24191 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :   EST:*
                1: em_estba:*
                2: em_esthum:*
                3: em_estin:*
                4: em_estmu:*
                5: em_estov:*
                6: em_estpl:*
                7: em_estro:*
                8: em_htc:*
                9: gb_est1:*
                10: gb_est2:*
                11: gb_htc:*
                12: gb_est3:*
                13: gb_est4:*
                14: gb_est5:*
                15: em_estfun:*
                16: em_estom:*
                17: em_gss_hum:*
                18: em_gss_inv:*
                19: em_gss_pln:*
                20: em_gss_vrt:*
                21: em_gss_fun:*
                22: em_gss_mam:*
                23: em_gss_mus:*
                24: em_gss_pro:*
                25: em_gss_rod:*
                26: em_gss_phg:*
                27: em_gss_vrl:*
                28: gb_gss1:*
                29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Query Length	DB	ID	Description
1	1501	97.1	1201	13	BX394121	BX394121 BX394121
2	1498	96.9	1201	9	AL560874	AL560874 AL560874
3	1455	94.1	1054	12	BM541420	BM541420 AGENCOURT
4	1446	93.5	1060	12	BM810651	BM810651 AGENCOURT
5	1432	92.6	952	9	AL554334	AL554334 AL554334
6	1415.5	91.6	944	12	BG829086	BG829086 602752428
7	1406	90.9	1196	13	BX416456	BX416456 BX416456
8	1345	87.0	893	13	BU175613	BU175613 AGENCOURT
9	1340	86.7	949	12	BG831321	BG831321 602766160
10	1288	83.3	765	14	CA775004	CA775004 ip11g06.y
11	1281	82.9	935	13	BQ921568	BQ921568 AGENCOURT
12	1239.5	80.2	779	10	BE746490	BE746490 601579445
13	1229	79.5	818	13	BU595053	BU595053 AGENCOURT
14	1222	79.0	916	12	BM451915	BM451915 AGENCOURT
15	1215	78.6	869	10	BF984400	BF984400 602307923
16	1210	78.3	707	12	BI223765	BI223765 602943149
17	1200	77.6	959	10	BF689462	BF689462 602186827
18	1198.5	77.5	714	10	BE778327	BE778327 601463720
19	1195	77.3	946	10	BF038506	BF038506 601460421
20	1155	74.7	764	12	BG767598	BG767598 602741757
21	1153	74.6	1198	12	BM805842	BM805842 AGENCOURT
22	1110.5	71.8	693	10	BG768406	BG768406 602742157
23	1100	71.2	864	10	BG420594	BG420594 602448473
24	1090.5	70.5	782	13	BU929741	BU929741 AGENCOURT
25	1088.5	70.5	772	12	CB961021	CB961021 AGENCOURT
26	1083.5	70.1	792	10	BE407918	BE407918 601299260
27	1067	69.0	613	12	BM782115	BM782115 K-EST0058
28	1058	68.4	652	12	BI518784	BI518784 603061972
29	1058	68.4	686	12	BG824728	BG824728 602728713
30	1052.5	68.1	937	10	BE786259	BE786259 601474394
31	1040.5	67.3	704	10	BF129039	BF129039 601811367
32	1018	65.8	1209	12	BM454170	BM454170 AGENCOURT
33	1003.5	64.9	868	13	BQ942688	BQ942688 AGENCOURT
34	1002	64.8	783	11	CB529180	CB529180 UI-H-FT2-
35	996	64.4	1364	11	AK011680	AK011680 Mus muscu
36	992	64.2	927	13	BI558492	BI558492 603240375
37	973	62.9	823	13	BQ896738	BQ896738 AGENCOURT
38	968.5	62.6	1102	12	BI410112	BI410112 602964039
39	967	62.5	619	10	BG719686	BG719686 602688896
40	960.5	62.1	718	12	BG778946	BG778946 602667481
41	952	61.6	801	10	BF690400	BF690400 602186827
42	949	61.4	534	12	BM839989	BM839989 K-EST0116
43	945	61.1	807	9	AW006811	AW006811 wt07f11.x
44	944	61.1	775	14	CA435275	CA435275 UI-H-DP0-
45	942	60.9	574	10	AW965779	AW965779 EST377852

ALIGNMENTS

RESULT 1
BX394121
LOCUS   BX394121   1201 bp   mRNA   linear   EST 13-MAY-2003
DEFINITION   BX394121 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens
               cDNA clone CS0DC010YI23 5-PRIME, mRNA sequence.
ACCESSION   BX394121
VERSION   BX394121.1   GI:30612409
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM   Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1201)

AUTHORS　　Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE　　Full-length cDNA libraries and normalization
JOURNAL　Unpublished
COMMENT　Contact: Genoscope
　　　　　Genoscope - Centre National de Sequencage
　　　　　BP 191 91006 EVRY cedex - France
　　　　　Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
　　　　　Library was constructed by Life Technologies, a division of
　　　　　Invitrogen. This sequence belongs to sequence cluster 1105.r For
　　　　　more information about this cluster, see
　　　　　http://www.genoscope.cns.fr/
　　　　　cgi-bin/cluster.cgi?seq=CS0DC010AB12QP1&cluster=1105.r. Contact :
　　　　　Feng Liang Email : fliang@lifetech.com URL :
　　　　　http://fulllength.invitrogen.com/ InvitroGen Corporation 1600
　　　　　Faraday Avenue Genoscope sequence ID : CS0DC010AB12QP1.
FEATURES　　　　　　　Location/Qualifiers
　　　　source　　　1..1201
　　　　　　　　　　/organism="Homo sapiens"
　　　　　　　　　　/mol_type="mRNA"
　　　　　　　　　　/db_xref="taxon:9606"
　　　　　　　　　　/clone="CS0DC010YI23"
　　　　　　　　　　/tissue_type="NEUROBLASTOMA COT 25-NORMALIZED"
　　　　　　　　　　/clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
　　　　　　　　　　/note="1st strand cDNA was primed with a Noti-oligo(dT)
　　　　　　　　　　primer. Five prime end enriched, double-strand cDNA was
　　　　　　　　　　digested with Not I and cloned into the Not I and EcoR V
　　　　　　　　　　sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT　　305 a　　241 c　　340 g　　275 t　　40 others
ORIGIN

Alignment Scores:
Pred. No.:
Score:　　　　　　　1.51e-136
　　　　　　　　　　1501.00
Percent Similarity:　99.30%
Best Local Similarity:　98.25%
Query Match:　　　　97.09%
DB:

Length:　　　1201
Matches:　　281
Conservative:　3
Mismatches:　2
Indels:　　　1
Gaps:　　　0

US-10-054-313-1 (1-286) x BX394121 (1-1201)

Qy　　1　MetSerTrpLeuPheLeuAlaHisArgValAlaAlaAlaLeuAlaAlaProCysArgArg　20
　　　　　|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　　99　YTGAGCTGGCTTCGTTCCTGGCCCACAGAGTCGCCTTGGCCGCCTGCCCTGCCGCCGC　158

Qy　　21　GlySerArgGlyPheGlyMetPheTyrAlaValAlaArgGlyArgLysThrGlyValPhe　40
　　　　　|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　　159　GGCTCTTCGCCGGTTCGGATGTTCTATGCCGTGAGGAGGGCCCGAAAGACCGGGGTCTTT　218

Qy　　41　LeuThrTrpAsnGluCysArgAlaGlnAsnValAspArgPheProAlaAlaArgPheLysLys　60
　　　　　|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　　219　CTGACCTGGAATGAGTGCAGAGCACAGGTGGACCCGGTTTCTGCTGCCAGATTTAAGAAG　278

Qy　　61　PheAlaThrGluGluAspGluAlaTrpAlaTrpAlaValArgGlyLysSerAlaSerProGluValSer　80
　　　　　|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　　279　TTTGCCACAGAGGATGAGGGCCTGGGCTTTGTCAGGAAATCTGCAAGCCCGGAAGTTTCA　338

Qy　　81　GluGlyHisGluAsnGlnHisGlyIleGlnGluSerGluAlaLysProGlyLysArgLeuArg　100
　　　　　|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　　339　GAAGGGCATGAGAAAATCAACATGGACAAGAATCGGAGGGCGAAAGT-AGCAAGCGACTCCGT　397

Qy　　101　GluProLeuAspGlyHisAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro　120
　　　　　|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　　398　GAGCCACTGGATGAGGATGGACATGGAAGCGCCAGACCGTATGCAAAGCACATGAAGCCG　457

Qy　　121　SerValGluProAlaProProValSerArgArgAspThrProSerTyrMetGlyAspPheVal　140
　　　　　|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　　458　AGCGTGGAGCCGGCCGCCCTCCAGTTAGCAGAGACACGTTTTCCTACATGGGAGACTTCGTC　517

Qy　　141　ValValTyrThrArgGlyTyrCysCysSerSerArgAsnGlyArgArgLysProArgArgAlaGlyIle　160
　　　　　||||||||||||||||||||||||||||::::|||||||||||||||||||||||||||
Db　　518　GTCGTCTACACTCGATGGCTGCTGCTCCAGTAATGGGCGTAGAAGGCCGAGCAGGAATC　577

Qy　　161　GlyValTyrTrpProGlyHisProLeuAsnValGlyIleArgGlyProGlyArgGln　180

Db　　578　GGCGGTTACTGGGGGCAGGCCATCCTTTAAATGTAGGCATTAGGACTTCCTGGGGGCAG　637
　　　　　|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy　　181　ThrAsnGlnArgGlyAlaGluIleHisAlaAlaCysLysLysAlaIleGluGlnAlaLysThrGln　200

Db　　638　ACAAACCAAGAGCGGAAATTCATGCAGCCTGCAAAGCCATTGAACAAGCAAAGACTCAA　697
　　　　　|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy　　201　AsnIleAsnLysLeuValLeuTyrThrArgAspSerMetPheThrIleArgGlyIleThrAsn　220

Db　　698　AACATCAATAAACTGGTTCTGTATACAGATAGCAGTATGTTTACGATAAATGGTATAACTAAC　757
　　　　　|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy　　221　TrpValGlnGlyTrpLysLysAsnGlyTyrTrpLysThrSerAlaGlyValLysGluValIleAsn　240

Db　　758　TGGGTTCAAGGTTGGAAGAAAAATGGGTGGAAGACAAGTGCAGGGAAAGAGGTGATCAAC　817
　　　　　|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy　　241　LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleTrpHis　260

Db　　818　AAAGAGACTTTGTGGCCACTGGAGAGAGCTTACCCAGGGATGGACATTCAGTGGATGCAT　877
　　　　　|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy　　261　ValProGlyHisSerGlyPheIleGlyValAsnGluAlaAspArgLeuAlaArgGluGly　280

Db　　878　GTTCCTGGTCATTCGGGATTTATAGGCAATGAAGAAGCTGCAGATTAGCCAGAGAAGGA　937
　　　　　|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy　　281　AlaLysGlnSerGluAsp　286

Db　　938　GCTAAACAATCGGAAGAC　955

RESULT 2
AL560874
LOCUS　　　　AL560874　　　　　　1201 bp　　mRNA　　linear　　EST 31-MAY-2003
DEFINITION　AL560874 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED
　　　　　　Homo sapiens cDNA clone CS0DL005YJ14 5-PRIME, mRNA sequence.
ACCESSION　AL560874
VERSION　　AL560874.2　GI:31285003
KEYWORDS　EST.
SOURCE　　Homo sapiens (human)
ORGANISM　Homo sapiens
　　　　　Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
　　　　　Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE　1　(bases 1 to 1201)
AUTHORS　　Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE　　Full-length cDNA libraries and normalization
JOURNAL　Unpublished
COMMENT　On Feb 15, 2001 this sequence version replaced gi:12907756.
　　　　　Contact: Genoscope
　　　　　Genoscope - Centre National de Sequencage
　　　　　BP 191 91006 EVRY cedex - France
　　　　　Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
　　　　　Library was constructed by Life Technologies, a division of
　　　　　Invitrogen. This sequence belongs to sequence cluster 1105.r For
　　　　　more information about this cluster, see
　　　　　http://www.genoscope.cns.fr/
　　　　　cgi-bin/cluster.cgi?seq=CS0DL005DE07QP1&cluster=1105.r. Contact :
　　　　　Feng Liang Email : fliang@lifetech.com URL :
　　　　　http://fulllength.invitrogen.com/ InvitroGen Corporation 1600
　　　　　Faraday Avenue Genoscope sequence ID : CS0DL005DE07QP1.
FEATURES　　　　　Location/Qualifiers
　　　source　　1..1201
　　　　　　　　/organism="Homo sapiens"
　　　　　　　　/mol_type="mRNA"
　　　　　　　　/db_xref="taxon:9606"
　　　　　　　　/clone="CS0DL005YJ14"
　　　　　　　　/cell_type="B CELLS (RAMOS CELL LINE)"
　　　　　　　　/cell_line="RAMOS CELL LINE"
　　　　　　　　/clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE) COT
　　　　　　　　25-NORMALIZED"
　　　　　　　　/note="1st strand cDNA was primed with a NotI-oligo(dT)
　　　　　　　　primer. Five prime end enriched, double-strand cDNA was
　　　　　　　　digested with Not I and cloned into the Not I and EcoR V
　　　　　　　　sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT　　300 a　　248 c　　358 g　　259 t　　36 others
ORIGIN

Right column

```
DEFINITION   AGENCOURT_6493820 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5521285
             5', mRNA sequence.
ACCESSION    BM541420
VERSION      BM541420.1  GI:18770049
KEYWORDS     EST.
SOURCE       Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 1054)
  AUTHORS    NIH-MGC http://mgc.nci.nih.gov/.
  TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL    Unpublished
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLAM12186 row: k column: 14
             High quality sequence stop: 713.
FEATURES             Location/Qualifiers
     source          1..1054
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5521285"
                     /tissue_type="leiomyosarcoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_71"
                     /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 2.1 kb.        "
BASE COUNT     274 a    239 c    323 g    214 t    4 others
ORIGIN
```

```
Alignment Scores:
Pred. No.:
Score:                4.03e-132
                      1455.00
Percent Similarity:   97.24%
Best Local Similarity: 95.86%
Query Match:          94.11%
DB:
Length:               1054
Matches:              278
Conservative:         4
Mismatches:           4
Indels:               4
Gaps:                 0

US-10-054-313-1 (1-286) x BM541420 (1-1054)

Qy    1 MetSerTrpLeuPheLeuAlaHisArgValAlaLeuAlaAlaLeuProCysArgArg  20
Db   45 ATGAGCTGGCTTCTGTTCCTGGCCCACAGAGTCGCCCTTGCCCGTGCCCGCCGC 104

Qy   21 GlySerArgGlyPheGlyMetPheTyrAlaValArgArgGlyLysThrGlyValPhe  40
Db  105 GGCTCTCGCGGGTTCGGATGTTCTATGCCGTGAGGAGGGGCCAAGACCGGGGTCTTT 164

Qy   41 LeuThrTrpAsnGluCysArgAlaGlnIleValAspArgProAlaAlaArgPheLysLys  60
Db  165 CTGACCTGGAATGAGTGCAGAGCACAGATTGGACCCGGTTCCTGCTGCCAGATTTAAGAAG 224

Qy   61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer  80
Db  225 TTTGCCACAGAGGATGAGGCCTGGGCCTTTGTCAGGAAAATCTGCAAGCCCGGAAGTTTCA 284

Qy   81 GluGlyHisGluAsnGlnHisGlyGlyGluSerAlaGlnIleAlaLysProGlyLeuArg 100
Db  285 GAAGGGCATGAAAATCAACATGGACAAGCATCGGAGGGGCAAAGCGAGCCAGCGACTCCGT 344

Qy  101 GluProLeuAspGlyHisGlyAspGlyHisGlySerAlaGlnIleProTyrAlaLysHisMetLysPro 120
Db  345 GAGCCACTGGATGGACATGGAGATGGACATGGACCAGAGCGCCAGAAGGCGCAGAGCCGTATGCAAAGCACATGAAGCCG 404
```

Left column

```
Alignment Scores:
Pred. No.:
Score:                2.96e-136
                      1498.00
Percent Similarity:   98.95%
Best Local Similarity: 98.25%
Query Match:          96.90%
DB:
Length:               1201
Matches:              281
Conservative:         2
Mismatches:           3
Indels:               1
Gaps:                 0

US-10-054-313-1 (1-286) x AL560874 (1-1201)

Qy    1 MetSerTrpLeuPheLeuAlaHisArgValAlaLeuAlaAlaLeuProCysArgArg  20
Db  142 ATGAGCTGGCTTCTGTTCCTGGCCCACAGAGTCGCCCTTGCCCGTGCCCGCCGC 201

Qy   21 GlySerArgGlyPheGlyMetPheTyrAlaValArgArgGlyLysThrGlyValPhe  40
Db  202 GGCTCTCGCGGGTTCGGATGTTCTATGCCGTGAGGAGGGGCCAAGACCGGGGTCTTT 261

Qy   41 LeuThrTrpAsnGluCysArgAlaGlnIleValAspArgProAlaAlaArgPheLysLys  60
Db  262 CTGACCTGGAATGAGTGCAGAGCACAGATTGGACCCGGTTCCTGCTGCCAGATTTAAGAAG 321

Qy   61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer  80
Db  322 TTTGCCACAGAGGATGAGGCCTGGGCCTTTGTCAGGAAAATCTGCAAGCCCGGAAGTTTCA 381

Qy   81 GluGlyHisGluAsnGlnHisGlyValGlyLeuSerGluAlaGlnIleAlaLysProGlyLysLeuArg 100
Db  382 GAAGGGCATGAAAATCAACATGGACAAGCATCGGAGGGGCAAAGC-AGCAAGCCCGAACTCCGT 440

Qy  101 GluProLeuAspGlyHisGlyAspGlyHisGlyGluSerAlaGlnIleProTyrAlaLysHisMetLysPro 120
Db  441 GAGCCACTGGATGGACATGGAGATGGACATGGAGAGTCAAAGCCAGAGCCGTATGCAAAGCACATGAAGCCG 500

Qy  121 SerValGluProAlaProProValSerArgAspAspAspArgPheSerTyrMetGlyMetGlyAspPheVal 140
Db  501 AGCGTGGAGCCGGCGCCGCCCGTCAGTAGCGAGACCACGTTTTCCCACATGGGATGGGAGACTTCGTC 560

Qy  141 ValValTyrThrAspTyrGlyCysCysSerSerAsnGlyTyrArgArgLysProArgArgAlaGlyIle 160
Db  561 GTCGTCTACACTGATGGCTGCTGCCCAGTAATGGGCGTAGAGGGCGGCAGCAGGAATC 620

Qy  161 GlyValTyrTrpGlyTyrProGlyHisProGluProAsnValGlyIleArgGlyLeuProGlyArgGln 180
Db  621 GGCGGTTACTGGGGCGCATCCTTTAAATGGAGCATTAGGACTTCCTGGGCGGCAG 680

Qy  181 ThrAsnGlnAsnArgAlaGluIleHisAlaAlaLeuCysLysLysAlaAlaIleGluGluGlnAlaLysLysThrGln 200
Db  681 ACAAACCAAAGCGGCGAAATTCATGCAGCCTGCAAAGCCTGCAATCGAACAGCAAAGACTCAA 740

Qy  201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220
Db  741 AACATCAATAAACTGGTTCTGTACACAGACAGTATGTTTACGATAAATGGTATAAACTAAC 800

Qy  221 TrpValGlnLeuTyrTrpLysLysAsnGlyTyrTrpLysTrpThrSerAlaGlyLysGluValIleAsn 240
Db  801 TGGGTTCAAGGTTGGAAGAAAAAATGGGTTGGAAGAAGTCGCAGGGAAAGGTGATCAAC 860

Qy  241 LysGluAspPheValAlaAlaGlyLeuThrGlnLeuArgLeuMetAspIleGlnTrpMetHis 260
Db  861 AAAGAGGACTTTGTGGCACTGGAGAGGCTTACCCAGGGGATGGACATTCAGTGGATGCAT 920

Qy  261 ValProGlyValHisSerGlyPheIleGlyTyrAsnGluIleAlaAspArgLeuAlaAspArgGluGluGly 280
Db  921 GTTCCTGGGTCATTGGGATTTATAGGCCAATGAAGGAGTGTGACAAATTAGCCAGGGAAGGA 980

Qy  281 AlaLysGlnSerGluAsp 286
Db  981 GCTAAACAATCGGAGAC 998
```

RESULT 3

BM541420

LOCUS   BM541420   1054 bp   mRNA   linear   EST 20-FEB-2002

EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

BASE COUNT   265 a   246 c   332 g   213 t   4 others
ORIGIN

Alignment Scores:
Pred. No.:   Length: 1060
Score: 3.09e-131   Matches: 274
Percent Similarity: 1446.00   Conservative: 4
Best Local Similarity: 96.86%   Mismatches: 8
Query Match: 95.47%   Indels: 2
DB: 93.53%   Gaps: 0
                            12

US-10-054-313-1 (1-286) x BM810651 (1-1060)

```
Qy    1  MetSerTrpLeuPheLeuLeuHisArgValAlaLeuLeuAlaAlaLeuProCysArgArg   20
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   58  ATGAGCTGGCTTCTGTTCCTGCCCCACAGCGTGGCCCTTGGCCCTGCCCGCCGC  117

Qy   21  GlySerArgGlyPheGlyMetPheTyrAlaValAlaArgGlyThrGlyValPhe   40
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  118  GGCTCTCGCGGGTTCGGGATGTTCTATGCCGTGAGGAGGCGCAAGACCGGGGTCTTT  177

Qy   41  LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys   60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  178  CTGACCTGGAATGAGTGCAGAGCACAGGTGGACCGGTTCCCTGCTGCCAGATTTAAGAAG  237

Qy   61  PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer   80
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  238  TTTGCCACAGAGGATGAGGCCTGGGCCTTTGTCAGGAAATCTGCAAGCCCGGAAGTTTCA  297

Qy   81  GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaGlyGlyLysArgLeuArg  100
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  298  GAAGGGCATGAAGAAATCAACATGGACAAGAAATCGAGGCGAAAGCGACTCCGT  357

Qy  101  GluProLeuAspGlyHisGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro  120
         |||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||||
Db  358  GAGCCACTGGATGGACATGGACATGGAGAGCGCAGAGCCGTATGCAAAGCACATGAAGCCG  417

Qy  121  SerValGluProAlaProProValSerArgAspThrPheSerTyrMetGlyAspPheVal  140
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  418  AGCGTGGAGCCGGCCCCTCCAGTAGCAGAGACACGTTTCTCATACTGGGAGACTTCGTC  477

Qy  141  ValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGlyIle  160
         |||||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||
Db  478  GTCGTCACACTGATGGCTGCTGCTCCAGTAATGGGCCGAGAAGGGCCGAGCCAGGAAATC  537

Qy  161  GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyIleArgArgLeuProGlyArgGln  180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  538  GGGTTTACTGGGGGCCCGGCCCATCCTTTAAATGTAGGCATTAGACTTCCTGGGCGGCAG  597

Qy  181  ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaAlaIleGluGlnAlaLysThrGln  200
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  598  ACAAACCAAAGAGCGGAAATTCATGCAGCCTGCAAAGCCATTGAACAAGCAAAGACTCAA  657

Qy  201  AsnIleAsnLysAsnValLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn  220
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  658  AACATCAATAAACTGGTTCTGTATCACAGATCAGTATGTTTACGATAAATGGTATAACTAAC  717

Qy  221  TrpValGlnGlyTrpLysAsnGlyTrpLysThrSerAlaGlyLysGluValIleAsn  240
         ||||||||||||||| |||::  ::|||||||||||||||||||||||||||:::
Db  718  TGGGTTCAAGGTTGGAAGAA-AATGGGTGAAGAA-AATGGGTGAAGAAGAGGTGATNCAC  776

Qy  241  LysGluAspPheValAlaLeuUThrGlnLeuThrGlnIleGlyMetAspIleGlnTrpGlyMetHis  260
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  777  AAAGAGGACTTTGTGCACTGGACTGGAGGCTTACCCAGGGATGGACATTCAGTGGATGCAT  836

Qy  261  ValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArg-LeuAlaArgGluGluGl  280
```

```
Qy  121  SerValGluProAlaProProValSerArgAspThrPheSerTyrMetGlyAspPheVal  140
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  405  AGCGTGGAGCCGGCCGCTCCAGTTAGCAGAGACACGTTTCTCCATACTGGGAGACTTCGTC  464

Qy  141  ValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGlyIle  160
         |||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||||
Db  465  GTCGTCACACTGATGGCTGCTGCTCCAGTAATGGGGCGTAGAAGGGCCGAGCCAGGAATC  524

Qy  161  GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyIleArgArgLeuProGlyArgGln  180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  525  GGGCCGTTTACTGGGGGCCCGGCCCATCCTTTAAATGTAGGCATTAGACTTCCTGGGCGGCAG  584

Qy  181  ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaAlaIleGluGlnAlaLysThrGln  200
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  585  ACAAACCAAAGAGCGGAAATTCATGCAGCCTGCAAAGCCATTGAACAAGCAAAGACTCAA  644

Qy  201  AsnIleAsnLysAsnValLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn  220
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  645  AACATCAATAAACTGGTTCTGTATCACAGATCAGTGTTTACGATAAATGGTATAACTAAC  704

Qy  221  TrpValGlnGlyTrpLys-LysAsnGlyTrpLysThrSerAla-GlyLysGluValIleA  240
         ||||||||||||||| |||:::|||||||::||||||||||||:::||||||||||||
Db  705  TGGGTTCAAGGTTGGAAGNAAAAATGGGTGGAAGACAAGTGCAGGGGAAAAGAGGTGATCA  764

Qy  240  snLysGluAspPheValAlaLeuUArgGluLeuThrGln-GlyMetAspIleGlnTrpMet  259
         ||||||||||||||||||||||||||||:||||||||||||||||||||||||||||||||
Db  765  ACAAAGAGGACTTTGTGCACTGGAGAGGCTTACCCAGGGGNATGGACATTCAGTGGATG  824

Qy  260  HisValPro-GlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgGl  279
         |||||||||||||||||||||||||||:::|||||||||||||||||||||||||||||
Db  825  CATGTTCCTGGGTCATTCGGGATTTATAGGCCATGAAGAAAACTGACAGATTAGCCAGGAA  884

Qy  279  uGlyAlaLysGlnSerGluGluAsp  286
         ::|||||||||||||||||||||||
Db  885  AGGAGCTAAACCATCGGAAGAC  906
```

RESULT 4
BM810651
LOCUS   BM810651   1060 bp   mRNA   linear   EST 05-MAR-2002
DEFINITION   AGENCOURT_6580549 NIH_MGC_98 Homo sapiens cDNA clone IMAGE:5453776
             5', mRNA sequence.
ACCESSION   BM810651
VERSION   BM810651.1   GI:19127474
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1060)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished
COMMENT   Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: ATCC
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLCM1947 row: n column: 17
             High quality sequence stop: 725.
FEATURES   Location/Qualifiers
     source   1..1060
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5453776"
                     /tissue_type="astrocytoma grade IV, cell line"
                     /lab_host="DH10B (phage resistant)"
                     /clone_lib="NIH_MGC_98"
                     /note="Organ: brain; Vector: pCTB7; Site_1: XhoI; Site_2:

```
Db   837 GGTCCTGTCATTCGGGATTTATANGGCATGAAGAAGCTGACAGATTTANCCAGAGAAGG 896
         ||||||          :::||||||||||       |||||||||||
Qy   280 yAlaLysGlnSerGluAsp 286
Db   897 AGCTNAACAATCGGAAGAC 915

RESULT 5
AL554334
LOCUS          AL554334                952 bp    mRNA    linear   EST 31-MAY-2003
DEFINITION     AL554334 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
               clone CSODI082YL10 5-PRIME, mRNA sequence.
ACCESSION      AL554334
VERSION        AL554334.2  GI:31276146
KEYWORDS       EST.
SOURCE         Homo sapiens (human)
ORGANISM       Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1  (bases 1 to 952)
AUTHORS        Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE          Full-length cDNA libraries and normalization
JOURNAL        Unpublished
COMMENT        On Feb 15, 2001 this sequence version replaced gi:12895013.
               Contact: Genoscope
               Genoscope - Centre National de Sequencage
               BP 191 91006 EVRY cedex - France
               Email: seqref@genoscope.cns.fr
               Seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
               Library was constructed by Life Technologies, a division of
               Invitrogen. This sequence belongs to sequence cluster 1105.r For
               more information about this cluster, see
               http://www.genoscope.cns.fr/
               cgi-bin/cluster.cgi?seq=CSODI082DF05QP1&cluster=1105.r. Contact :
               Feng Liang Email : fliang@lifetech.com URL:
               http://fullength.invitrogen.com/ InVitroGen Corporation 1600
               Faraday Avenue Genoscope sequence sequence ID : CSODI082DF05QP1.
FEATURES             Location/Qualifiers
     source          1..952
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CSODI082YL10"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="list strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with NotI and cloned into the NotI and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT    243 a    217 c    300 g    192 t
ORIGIN

Alignment Scores:
Pred. No.:               6.28e-130
Score:                   1433.00             Length:         952
Percent Similarity:      99.25%              Matches:        263
Best Local Similarity:   98.50%              Conservative:     2
Query Match:             92.63%              Mismatches:       2
DB:                      9                   Indels:           0
                                             Gaps:             0

US-10-054-313-1 (1-286) x AL554334 (1-952)

Qy     1 MetSerTrpLeuPheLeuAlaHisArgValAlaHisAlaAlaLeuProCysArgArg 20
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   152 ATGAGCTGGCTTCGTTCCTTGGCCCACAGAGTGCCCATGCAGCTGCCCTGCCGGCGC 211
Qy    21 GlySerArgGlyPheGlyMetPheTyrAlaValArgArgValLysThrGlyValPhe 40
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   212 GGCTCGCGGGGTTCGGGATGTTCTATGCCGTGAGGAGGGCGCAAGACCGGGTCTTT 271
Qy    41 LeuThrTrpAspAsnGluCysArgArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   272 CTGACCTGGAATGAGTGCAGAGCAGGTGGACCCGGTTTCCTGCTGCCAGATTTAAGAAG 331
```

```
Qy    61 PheAlaThrGluAspGluAlaTrpAlaPheValArgGlySerGluValSer 80
         ||||||||     |||||       :::|||||||
Db   332 TTTGCCACAGAGGATGAGGCCTGGGCCTTTGTCAGGAAATCTGAAGTTTCA 391
Qy    81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGlyAlaLysProGlyLysArgLeuArg 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   392 GAAGGGCATGAAAATCAACATGGACAAGAATCGGAGGGAAAGCCAGCAAGCGACTCCGT 451
Qy   101 GluProLeuAspGlyHisAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120
         ||||||||||||||||||||||||||||||||:::|||||||||
Db   452 GAGCCACTGGATGGAGATGGACATGAAAGCCGAGCCGTATGCAAAGCACATGAAGCCG 511
Qy   121 SerValGluProAlaProProValSerArgArgAspThrPheSerTyrMetGlyAspPheVal 140
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   512 AGCGTGGAGCCGGCGCCTCCAGTTAGCAGAGACACGTTTTCCTACATGGGAGACTTCGTC 571
Qy   141 ValValTyrThrAspGlyValCysCysSerSerAsnGlyArgArgLysProArgAlaGlyIle 160
         |||||||||||||||||||||||||||||||:::|||||||||
Db   572 GTCGTCTACACTGATGGCGTGCTGCTCCAGTAATGGCCGTGAGAAGGCCGGAGCAGGAATC 631
Qy   161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyTyrLeuArgLeuProGlyArgGln 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   632 GGCGTTTACTGGGGACCAGGCCATCCTTTAAATGTAGGCATTAGACTTCCTGGGCGGCAG 691
Qy   181 ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaIleGluGlnLnAlaLysThrGln 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   692 ACAAACCAAAGAGCGGAAATTCATGCAGCCTGCCAAGGCCATTGAACAAGCAAAGACTCAA 751
Qy   201 AsnIleAsnLysValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   752 AACATCAATAAACTGGTTCTGTATACAGACAGTATGTTTACGATAAATGTATAACTAAC 811
Qy   221 TrpValGlnAspGluLysLysAsnGlyTrpLysThrSerArgAlaGlyValIleAsn 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   812 TGGGTTCAAGGTTGGAAGAAAAATGGGTGGAAGACAAGTCCAGGGAAAGAGGTGGCAAC 871
Qy   241 LysGluAspPheValAlaLeuGluArgLeuThrGlnMetAspIleGlnTrpMetHis 260
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   872 AAAGAGGACTTTGTGGCACTGGAGAGGCTTACCCAGGGGATGGACATTCAGTGGATGCAT 931
Qy   261 ValProGlyHisSerGlyPhe 267
         ||||||||||||||||||||
Db   932 GTTCCTGGTCATTCGGGATTT 952

RESULT 6
BG829086
LOCUS          BG829086
DEFINITION     6027S2428F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4905246 5',
               mRNA sequence.
ACCESSION      BG829086
VERSION        BG829086.1  GI:14176673
KEYWORDS       EST.
SOURCE         Homo sapiens (human)
ORGANISM       Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1  (bases 1 to 944)
AUTHORS        NIH-MGC http://mgc.nci.nih.gov/.
TITLE          National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL        Unpublished
COMMENT        Contact: Robert Strausberg, Ph.D.
               Email: cgapbs-r@mail.nih.gov
               Tissue Procurement: ATCC
               cDNA Library Preparation: Ling Hong/Rubin Laboratory
               cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
               DNA Sequencing by: Incyte Genomics, Inc.
               Clone distribution: MGC clone distribution information can be
               found through the I.M.A.G.E. Consortium/LLNL at:
               http://image.llnl.gov
               Plate: LLCM1806  row: g   column: 07
               High quality sequence stop: 781.
FEATURES             Location/Qualifiers
     source          1..944
```

```
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4905246"
                     /tissue_type="rhabdomyosarcoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH MGC 17"
                     /note="Organ: muscle; Vector: pOTB7; Site_1: EcoRI;
                     Site_2: XhoI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Size-selected >500bp
                     for average insert size 1.8kb. Library constructed by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT    255 a    204 c    295 g    190 t
ORIGIN

Alignment Scores:
Pred. No.:
Score:                 2.56e-128          Length:        944
Percent Similarity:    1415.50            Matches:       271
Best Local Similarity: 96.18%             Conservative:  6
Query Match:           94.10%             Mismatches:    8
DB:                    91.56%             Indels:        4
                       12                 Gaps:          1

US-10-054-313-1 (1-286) x BG829086 (1-944)

Qy   1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaAlaLeuProCysArgArg  20
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4 ATGAGCTGGCTTCTTGTTCCTGGCCCACAGAGTCGCCTTGGCCGCCTTGCCCGTCGCCGC  63

Qy  21 GlySerArgGlyMetPheTyrArgAlaValArgArgGlyArgLysThrGlyValValPhe  40
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  64 GGCTTCCGCCGGTTCGGGGATGTTCTATGGCCGTGAGGAGGGCCCGCAAGACCGGGGTCTTT 123

Qy  41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys  60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 124 CTGACCTGGAATGAGTGCAGAGCACAGGTGGACCCGGTTTCCTGCTGCCAGATTTAAGAAG 183

Qy  61 PheAlaThrGluAspGluAlaTrpAlaTrpAlaPheValArgLysSerAlaSerProGluValSer  80
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 184 TTTGCCACAGAGGATGAGGCCTGGGCCTTTGTCAGGAAATCTGCAAGCCCGGAAGTTTCA 243

Qy  81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyGlyAlaLysArgLeuArg 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 244 GAAGGGCATGAAAATCAACATGGACAAGAATCGGAGCGGAAAGCCAGCAGCCGGACTCCGT 303

Qy 101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 304 GAGCCACTGGATGAGATGGACATGAAAGTGCAGAGCCCTATGGCAAAGCACATGAAGCCG 363

Qy 121 SerValGluProAlaProProValSerArgArgAspThrPheSerTyrMetGlyAspPheVal 140
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 364 AGCGTGGAGCCGGCGCCGCCTCCAGTTAGCAGAGACACGTTTTCCTACATGGGAGACTTCGTC 423

Qy 141 ValValTyrThrAspGlyTyrCysSerSerAsnGlyArgArgGlyProAlaArgAlaGlyIle 160
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 424 GTCGTCTACACTGATGGCTGCTGCTCCAGTAATGGGCGTAGAAGGCCGCAGCCAGGAATC 483

Qy 161 GlyValTyrTrpGlyTyrTrpGlyHisProLeuAsnValGlyIleLeuArgLeuProValArgGln 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 484 GGCGTTTACTGGGGCGCAGGCCATCCTTTAAATGTAGGCATTAGACTTCCTGGGCGGCAAG 543

Qy 181 ThrAsnGlnArgAlaGluHisHisAlaAlaAlaCys-LysAlaIleGluGluGlnAlaLysLysThrGl 200
         ||||||||||||||||||||||||||||||||||:||||||||||||||||||||||
Db 544 ACAAACCAAGAGGCCGAAAATTCATGCAGCCTGCAAAAGCCATTGAACAGCAAGACTCA 603

Qy 200 nAsnIleAsnLysLeuValLeuTyrThrAspGlySerMetPheThrIleAsnGlyIleThrArg 220
         |||||||||||||||||||||||||||||||||||||||:|||||||||||
Db 604 AAACATCAATAAACTGGTTCTGTATACAGACAGTATGTTTACGATAAATGGTATAACTAA 663

Qy 220 nTrpValGlnGlyTrpLysAsnGlyTrpLysSerThrSerAlaGlyLysGlyLysValIleAs 240
```

```
Db 664 CTGGGTTCAAGGTTGGAAGAAAAATGGGTGGAGACAAAAGTGCAGGCAAAGAGGTGATCAA 723

Qy 240 nLysGluAspPheValAla-LeuLeuArgLeuThrGlnGlyMetAspIleGlnTrpMetH 260
         |||||||||||||||||  |||||||||||:::|||||||||||||::||||||||||
Db 724 CAAAGAGGACTTTGTGCACTTGGAGAGGTTACCAAGGGGATGGACATTCAGTCGGATGC 783

Qy 260 isValProGlyHisSerGlyPheIleGlyValAsnGluGluAlaArgLeuAlaArgGluG 280
         |||||||||||||||||||||||||||||||||||:::||||||||||||||||||||
Db 784 ATGTACCGGGTCATCGGGATTAT---AGGCCATGAAGCAAGCTGACAGATTAGCCAGAGAAG 840

Qy 280 lyAlaLysGlnSerGluAsp 286
         ||||||||||||||:|||
Db 841 GAGTTAA-CAATCGGAAGAGAC 859


RESULT 7
BX416456
LOCUS       BX416456        1196 bp    mRNA    linear    EST 15-MAY-2003
DEFINITION  BX416456 Homo sapiens NEUROBLASTOMA Homo sapiens cDNA clone
            CSODA007YO12 5-PRIME, mRNA sequence.
ACCESSION   BX416456
VERSION     BX416456.1    GI:30763629
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1196)
AUTHORS     Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE       Full-length cDNA libraries and normalization
JOURNAL     Unpublished
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: segref@genoscope.cns.fr; Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 1105.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DA007BH06QP1&cluster=1105.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitrogen Corporation 1600
            Farday Avenue Genoscope sequence ID : CS0DA007BH06QP1.
FEATURES             Location/Qualifiers
     source          1..1196
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DA007YO12"
                     /tissue_type="NEUROBLASTOMA"
                     /clone_lib="Homo sapiens NEUROBLASTOMA"
                     /note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
                     with a NotI-oligo(dT) primer. Five prime end enriched,
                     double-strand cDNA was digested with Not I and cloned into
                     the Not I and EcoRV sites of the pCMVSPORT 6 vector.
                     Library was not normalized."
BASE COUNT    300 a    247 c    345 g    269 t    35 others
ORIGIN

Alignment Scores:
Pred. No.:
Score:                 3.02e-127          Length:        1196
Percent Similarity:    1406.00            Matches:       277
Best Local Similarity: 97.55%             Conservative:  2
Query Match:           96.85%             Mismatches:    7
DB:                    90.94%             Indels:        4
                       13                 Gaps:          0

US-10-054-313-1 (1-286) x BX416456 (1-1196)

Qy   1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaAlaLeuProCysArgArg  20
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 125 ATGAGCTGGCTTCTTGTTCCTGGCCCACAGAGTCGCCTTGGCCGCCTTGCCCGTCGCCGC 184
```

```
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTP/Gazdar
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13196   row: h   column: 24
High quality sequence stop: 672.

FEATURES        Location/Qualifiers
     source     1..893
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:6009671"
                /tissue_type="large cell carcinoma"
                /lab_host="DH10B (phage-resistant)"
                /clone_lib="NIH_MGC_68"
                /note="Organ: lung; Vector: pCMV-SPORT6; Site 1: NotI;
                Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                Average insert size 1.8 kb. Library constructed by Life
                Technologies."

BASE COUNT     237 a    197 c    279 g    178 t    2 others
ORIGIN

Alignment Scores:
Pred. No.:
Score:                    1.91e-121        Length:           893
                          1345.00          Matches:          270
Percent Similarity:       95.14%           Conservative:     4
Best Local Similarity:    93.75%           Mismatches:       4
Query Match:              87.00%           Indels:           10
DB:                       13               Gaps:             0

US-10-054-313-1 (1-286) x BU175613 (1-893)

Qy   1 MetSerTrpLeuPheLeuAlaHisArgValAlaAlaLeuAlaAlaLeuProCysArgArg 20
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  27 ATGAGCTGGCTTCTGTTCCTGCCCCACAGAGTCGCCTTGGCCGCCCTTGCCCTGCCGCGC 86

Qy  21 GlySerArgGlyPheGlyMetPheTyrAlaAlaValAlaArgGlyArgLysThrGlyValPhe 40
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  87 GGCTCTCGCGGGTTCGGGATGTTCATGCCGTAGCAGCAGGTGGCCCGAAGACCGGGGTCTTT 146

Qy  41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgProAlaAlaArgPheLysLys 60
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 147 CTGACCTGGAATGAGTGCAGAGCACAGGTGGACCGGCCGGTTCCTGCTGCCAGATTTAAGAAG 206

Qy  61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 207 TTTGCCACAGAGGATGAGGCCTGGGCCTTTGTCAGGAAATCTGCAAGCCCCGAAGTTTCA 266

Qy  81 GluGlyHisGluAsnGlnHisGlyGluSerGlyGlnGluGluSerAlaGlnProGlyLysSerProGlyLeuArg 100
       |||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db 267 GAAGGGCATGAAAATCAACATGGACAAGAATCGGCGAAAGCAGCCAGCAAGCGACTTCCGT 326

Qy 101 GluProLeuAspGlyAspHisGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120
       |||||||||||||||||||||||||||||||||||||||||||||||||||:||||
Db 327 GAGCCACTGGATGGAGATGAGAGCGCCCAGCCGTATGCAAAGCACATGAAGCCG 386

Qy 121 SerValGluProAlaProProValSerSerSerAsnGlyArgArgPheTyrMetGlyAspPheVal 140
       |||||||||||||||||||||||||:|||||||||||||||||||||||||:||||
Db 387 AGCGTGGAGCCGGCCGCCTCCAGTAGCAGAGACACGTTTCCTACATGGGAGACTTCGTC 446

Qy 141 ValValTyrThrArgThrSerSerSerAsnGlyArgArgLysProAlaArgAlaGlyIle 160
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||:||||
Db 447 GTCGTCTACACTGATGCTGGTGCTCGTCCAGTAATGACGTAATGGGCCGTAGAAGGCCGCAGGCAGGAAATC 506

Qy 161 GlyValTyrTrpGlyTyrProGlyHisProLeuLysArgLeuArgArgLeuIleArgGlyLeuArgGln 180
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||:||||
Db 507 GGCGTTTACTGGGGCCCGGGCCATCCTTAAATGTAGGCATTAGACTTCCTGGGCGCCAG 566

Qy 181 ThrAsnGlnArgAlaGluIleHisAlaAlaAlaCysLysLysAlaAlaIleGluGluGlnLysThrGln 200
```

```
Qy  21 GlySerArgGlyPheGlyMetPheTyrAlaAlaValAlaArgGlyArgLysThrGlyValPhe 40
Db 185 GGCTCTCGCGGGTTCGGGATGTTCATGCCGTAGCAGCAGGTGGCCCGAAGACCGGGGTCTTT 244

Qy  41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgProAlaAlaArgPheLysLys 60
Db 245 CTGACCTGGAATGAGTGCAGAGCACAGGTGGACCGGCCGGTTCCTGCTGCCAGATTTAAGAAG 304

Qy  61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80
Db 305 TTTGCCACAGAGGATGAGGCCTGGGCCTTTGTCAGGAAATCTGCAAGCCCCGAAGTTTCA 364

Qy  81 GluGlyHisGluAsnGlnHisGlyGlyTyrHisGlySerGluAlaLysProGlyLysArgLeuArg 100
Db 365 GAAGGGCATGAAAATCAACATGGACAAGAATCGGCGAAAGC-AGCMAGCCGACTCCGT 423

Qy 101 GluProLeuAspGlyAspHisGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120
Db 424 GAGCCACTGGATGGAGATGAGAGCGCCCAGCCGTATGCAAAGCACATGAAGCCG 483

Qy 121 SerValGluProAlaProProValSerSerSerAsnGlyArgArgPheTyrMetGlyAspPheVal 140
Db 484 AGCGTGGAGCCGGCCGCCTCCAGTTAGCAGAGACACGTTTCCTACATGGGAGACTTCGTC 543

Qy 141 ValValTyrThrAspGlyValCysCysSerSerAsnGlyArgArgLysProAlaArgAlaGlyIle 160
Db 544 GTCGTCTACACTGATGGCTGCTGCTCTCCAGTAATGGGCGTAGAGGCCGCGAGCAGGAAATC 603

Qy 161 GlyValTyrTrpGlyTyrProGlyHisProLeuAsnValGlyIleAsnValGlyIleTyrArgGln 180
Db 604 GGGGTTTACTGGGGCCCAGGCCATCCTTAAATGTAGCATTAGACTTCCTGGGCGCAG 663

Qy 181 ThrAsnGlnArgAlaGluIleHisAlaAlaAlaCysLysLysAlaAlaIleGluGluGlnLysThrGln 200
Db 664 AMAAACCAAAGAGCGAAATTCATGCAGCCTGCAAAGCCATTGAACAAGCAAAGACTCAA 723

Qy 201 AsnIleAsnLysLeuValLeuTyrThrArgSerMetPheThrIleAsnGlyIleThrAsn 220
Db 724 AACATCAATAAACTGGTTCTGTATACAAGACACGTATGTTTACGATAAATGGTATAACTAAC 783

Qy 221 TrpValGlnGlyTyrTrpLysTyrLysSerThrSerAlaGlyIleLysGluValIleLeuAsn 240
Db 784 TGGGTTCAAGGTTGGAAGAA-AATGGGTGGAAGACAAGTGCAGGGAAAGAGGTGATCAC- 841

Qy 241 LysGluAspPhePheValAlaAlaLeuArgGlyMetGlyMetAspIleGlnTrpMetHis 260
Db 842 AAAGAGGACTTTGTGGCACTGGAGAGGGCTTACCCAGGGGGATGGACATTCAGTGGATGCAT 901

Qy 261 ValProGlyHisSerGlyPheIleGlyValHisSerGlyGluGluAlaAspArgGluLeuAlaArgGluGly 280
Db 902 GTYCT-GGTCATTCGGGATTTATAGGCCATGAGGAGAGCTGACAGATTAGCCAGAGAAGGA 960

Qy 281 AlaLysGlnSerGluAsp 286
Db 961 GCTAAACAATCGGAAGAC 978
```

```
RESULT 8
BU175613
LOCUS       BU175613    893 bp    mRNA    linear    EST 04-SEP-2002
DEFINITION  AGENCOURT_7935024 NIH_MGC_68 Homo sapiens cDNA clone IMAGE:6009671
            5', mRNA sequence.
ACCESSION   BU175613
VERSION     BU175613.1    GI:22689597
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 893)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
```

Score: 1340.00   Matches: 257
Percent Similarity: 93.93%   Conservative: 6
Best Local Similarity: 91.79%   Mismatches: 15
Query Match: 86.68%   Indels: 3
DB: Gaps: 0

US-10-054-313-1 (1-286) x BG831321 (1-949)

```
Qy     1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaAlaLeuProCysArgArg  20
Db    30 ATGAGCTGGCTTCTGTTCCTGGCCCACAGAGTCGCCTTGGCCGCCCTTGCCCTGCCGCCGC  89

Qy    21 GlySerArgGlyPheGlyMetPheTyrAlaValArgArgGlyArgLysThrGlyValPhe  40
Db    90 GGCTCCCGCGGGTTCGGGATGTTCTATGCCGTGAGGAGGGGCCGCAAGACCGGGGTCTTT 149

Qy    41 LeuThrTrpAsnGluCysArgAlaGlnValAlaAspArgPheProAlaAlaArgPheLysLys  60
Db   150 CTGACCTGGAATGAGTGCAGAGCACAGGTGACCGGTTTCCTGCTGCCAGATTTAAGAAG 209

Qy    61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer  80
Db   210 TTTGCCACAGAGGATGAGGCCTGGGCCTTTGTCAGGAAATCTGCAAGCCCGGAAGTTTCA 269

Qy    81 GluGlyHisGluAsnGlnHisGlyTyrGlnGluSerGluAlaLysProGlyLysArgLeuArg 100
Db   270 GAAGGGCATGAAAATCAACATGGACAAGAAATCGGAGGCGAAAGCCAAGCGACTCCGT 329

Qy   101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120
Db   330 GAGCCACTGGATGGAGATGGACATGAAAGCCCAGACCCGATGCAAAGCACATGAAGCCG 389

Qy   121 SerValGluProAlaProProValSerArgArgAspThrPheSerTyrMetGlyAspPheVal 140
Db   390 AGCGTGGAGCCCGGCGCCTCCAGTTAGCAGAGACACGTTTCTCACATGGGAGACTTCGTC 449

Qy   141 ValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGlyIle 160
Db   450 GTCGTCTACACTGATGGCTGCTGCTCCAGTAATGGGCGTAGAAGGGCCGGAGCAGGAATC 509

Qy   161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyIleLeuArgGlyProGlyArgGln 180
Db   510 GGCGTTTACTGGGGACCCGGGCCATCCTTAAATGTAGGCATTAGACTTCCTGGGCGGCAG 569

Qy   181 ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaIleGluGlnAlaLysThrGln 200
Db   570 ACAAACCAAGAGAGCGGAAATTCATGCAGCCTGCAAAGCAATTGACAGCAAAGACTCAA 629

Qy   201 AsnIleAsnLysValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220
Db   630 AACATCAATAAACTGGTTCTGTATACAGACAGTATGTTTACGATAAATGGTATAACTAAC 689

Qy   221 TrpValGlnGlyTyr-LysLysAsn-GlyThrLysThrSerAlaGlyLysLysGluValIleAla 240
Db   690 TGGGTTCAAGGTTGGGAAGAAAAATGGGTTGGAAGACAAGTGCAGGCAAAGAGGTGATC 749

Qy   240 snLysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetH 260
Db   750 AAAAAGGAGGACTTTGTGGGACTCGAGAGGTT-ACCAGGGATGGGACATTCAGTGATGC 808

Qy   260 isValProGlyHisSerGlyPheIleGlyPheAsnGluGlnAlaAspArgLeuAlaArg 278
Db   809 ATGTCCTGGACATCCGGAATTATAGGAATGAAGGAAGTGAAGGATTAGCGAGA 864
```

RESULT 9
BG831321
LOCUS   BG831321   949 bp   mRNA   linear   EST 22-MAY-2001
DEFINITION   602766160F1 NIH_MGC_42 Homo sapiens cDNA clone IMAGE:4908326 5',
              mRNA sequence.
ACCESSION   BG831321
VERSION   BG831321.1 GI:14178908
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM   Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 949)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT   Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-r@mail.nih.gov
              Tissue Procurement: ATCC
              cDNA Library Preparation: Ling Hong/Rubin Laboratory
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: LLCM1814 row: g column: 15
              High quality sequence stop: 807.
FEATURES   Location/Qualifiers
     source   1..949
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4908326"
                     /tissue_type="epithelioid carcinoma cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_42"
                     /note="Organ: pancreas; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Size-selected >500bp
                     for average insert size 1.8kb. Library constructed by Ling
                     Hong in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
                     Note: this is a NIH_MGC Library.|"
BASE COUNT   255 a   195 c   303 g   196 t
ORIGIN

Alignment Scores: 6.4e-121   Length: 949
Pred. No.:

```
Db   567 ACAAACCAAGAGACGGAAATTCATGCAGCCTGCAAAGCCATTGACAGCAAAGACTCAA 626
Qy   201 AsnIleAsnLysValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220
Db   627 AACATCAATAAACTGGTTCTGTATACAGACAGTATGTTTACGATAAATGGTATAACTAAC 686

Qy   221 TrpValGlnGlyTyrLysLysAsnGlyTrpLysThrSerAlaGlyLysLysGluValIleAsn 240
Db   687 TGGGTTCAAGGTTGGGAAGAAAAATGGGTTGGAAGACAAGTGCAGGGAAAGAGGTGATCAAC 746

Qy   241 LysGluAspPhe-ValAlaLeu-GluArgLeuThrGlnGly--MetAspIleGlnTrp-M 259
Db   747 AAAGAGGACTTTGGTGGTGGCACTGGAGAAGCTTACCAGGGGGAATGGGACATTCAGTGGGA 806

Qy   259 etHis-ValProGly-His-SerGlyPheIle-GlyAsn-GluGluAlaAspArgLeuAl 277
Db   807 TGCATGGTTCCTGGNCCATTTCGGGAATTTATAGGGCAATGGAAGAAACTGACAGATTAGC 866

Qy   277 aArg 278
Db   867 CCAG 870
```

RESULT 10
CA775004
LOCUS   CA775004   765 bp   mRNA   linear   EST 03-DEC-2002
DEFINITION   ip1ig06.y1 HR85 islet Homo sapiens cDNA clone IMAGB:6216755 5'
              similar to TR:060523 060523 RIBONUCLEASE H TYPE II. [2] TR:060857
              ;, mRNA sequence.
ACCESSION   CA775004
VERSION   CA775004.1 GI:26012477
KEYWORDS   EST.

SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 765)
AUTHORS   Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
            Lemishka,I., Scearce,M., Breseteli,J., Gradwohl,G., Clifton,S.,
            Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
            Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas
            ,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T.
            , Jackson,Y. and Bowers,Y.
TITLE   Endocrine Pancreas Consortium
JOURNAL   Unpublished
COMMENT   Other_ESTs: ipl1g06.x1
            Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
            Endocrine Pancreas Consortium
            Harvard University, Howard Hughes Medical Institute
            Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
            MA 02138
            Tel: 617-495-1812
            Fax: 617-495-8557
            Email: dmelton@biohp.harvard.edu
            Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:
            Washington University Genome Sequencing Center For information on
            obtaining a clone please contact: Dr. Hiroshi Inoue
            (hinoue@im.wustl.edu)
            Seq primer: -40RP from Gibco
            High quality sequence stop: 469.
FEATURES   Location/Qualifiers
     source   1..765
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6216755"
                     /tissue_type="Purified pancreatic islet"
                     /lab_host="DH10B"
                     /clone_lib="HR85 islet"
                     /note="Organ: Pancreas; Vector: pBluescript SK(-); Site_1:
                     NotI; Site_2: XhoI; cDNA made by oligo-dT priming.
                     Size-selected on agarose gel. Average insert size ~1kb. 5'
                     XhoI site was destroyed after directional cloning.
                     Amplified once. Contact information: Hiroshi Inoue, MD,
                     Metabolism Div. (Alan Permutt Lab), Washington University
                     School of Medicine, Box 8127, 660 South Euclid Ave., St.
                     Louis, MO 63110, E-mail: hinoue@imgate.wustl.edu, Tel:
                     314-362-1916, Fax: 314-747-2692."
BASE COUNT   202 a   177 c   237 g   149 t
ORIGIN

Alignment Scores:
Pred. No.:   Length: 765
Score: 5.91e-116   Matches: 237
Percent Similarity: 1288.00   Conservative: 3
Best Local Similarity: 99.17%   Mismatches: 2
Query Match: 97.93%   Indels: 0
DB: 83.31%   Gaps: 0
                   14

US-10-054-313-1 (1-286) x CA775004 (1-765)

Qy   1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaAlaLeuProCysArgArg 20
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   40 ATGAGCTGGCTTCTGTTCCTGGCCCACAGAGTGGCCCTGGCCGCCCTTGCCCTGCCGCGC 99

Qy   21 GlySerArgGlyPheGlyMetPheTyrAlaValArgArgGlyArgLysThrGlyValPhe 40
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 100 GGCTCTCGCGGGTTCGGAGTGTTCTATGCCGTGAGGAGGGGCCGCAAGACCGGGGTCTTT 159

Qy   41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 160 CTGACCTGGAATGAGTGCAGAGCACAGGTGGACCGGTTTCCTGCTGCTCGCAGATTACAAG 219

Qy   61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80
        ||||||||||||||||||||||||||||||||||||

Db 220 TTTGCCACAGAGGATGAGGCCTGGGCCTTTGTCAGGAAATCTGCAAGCCCGGAAGTTTCA 279

Qy   81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGlnAlaLysProGlyLysArgLeuArg 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 280 GAAGGGCATGAAAATCAACATGGACAAGAATCGGAGGCGAAAGCCAGCAAGCGACTCCGT 339

Qy 101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120
        ||||||||||||||||||||||||||||||||||| :::||||||||||||||||||||||
Db 340 GAGCCACTGGATGGAGATGGACATGAAAGCCAGAGCCGTATGCAAAGCACATGAGAGCCG 399

Qy 121 SerValGluProAlaProProValSerArgAspThrProSerTyrMetGlyAspPheVal 140
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 400 AGCGGTGGAGCCGGCCGCCTCCAGTTAGCAGAGACACGTTTTCTACATGGGAGACTTCGTC 459

Qy 141 ValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGlyIle 160
        |||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db 460 GTCGTCTACACTGATGGCTGCTGCTCCAGTAATGGCCGTAGAAGGCCGCGAGCCAGGAATC 519

Qy 161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyIleLeuProGlyArgGln 180
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 520 GGCGGTTTACTGGGGCCCAGGCCATCTTTAAATGTAGGCATTAGACTTCCTGGGGCCAG 579

Qy 181 ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaIleGluGlnAlaLysThrGln 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 580 ACAAACCAAAGAGCGGAAATTCATGCAGCCTGCAAAGCCATTGAACAAGCAAAGACTCAA 639

Qy 201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 640 AACACTCAATAAACTGGTTCTGTATACACAGCAGTATGTTTACGATAAATGGTATAACTAAC 699

Qy 221 TrpValGlnGlyTrpLysAsnGlyTrpLysThrSerAlaGlyLysGluValValIleAsn 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||| :::
Db 700 TGGGTTCAAGGTTGGAAGAAAATGGTGGAAGACAAGTGCAGGGGAAAGAGTGATCCAC 759

Qy 241 LysGlu 242
        ||||||
Db 760 AAAGAG 765

RESULT 11
BQ921568
LOCUS   BQ921568   935 bp   mRNA   linear   EST 20-AUG-2002
DEFINITION AGENCOURT_8922346 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:6470333
            5', mRNA sequence.
ACCESSION   BQ921568
VERSION   BQ921568.1 GI:22336599
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 935)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM14000 row: o column: 06
            High quality sequence stop: 563.
FEATURES   Location/Qualifiers
     source   1..935
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6470333"
                     /tissue_type="leiomyosarcoma"
                     /lab_host="DH10B (phage-resistant)"

```
RESULT 12
LOCUS       BE746490
DEFINITION  BE746490
            601579445F1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3928402 5',
            mRNA sequence.
ACCESSION   BE746490
VERSION     BE746490.1  GI:10160482
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 779)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbb-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM759 row: i column: 11
            High quality sequence stop: 746.
FEATURES             Location/Qualifiers
     source          1..779
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3928402"
                     /tissue_type="adenocarcinoma cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_9"
                     /note="Organ: ovary; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G). Size-selected >500bp for average
                     insert size 1.8kb. Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT    218 a    161 c    242 g    158 t
ORIGIN
```

Alignment Scores:
```
Pred. No.:              3.41e-111        Length:          779
Score:                  1239.50          Matches:         238
Percent Similarity:     95.63%           Conservative:    3
Best Local Similarity:  94.44%           Mismatches:      8
Query Match:            80.17%           Indels:          4
DB:                                      Gaps:            1
```

US-10-054-313-1 (1-286) x BE746490 (1-779)

```
Qy   18 CysArgArgGlySerArgGlyPheGlyMetPheTyrAlaAlaValArgArgGlyArgLysThr  37
         |||||                    |||  ||||  || |        |||  ||| ||||
Db    2 TGCCGCGGG---TCTCGCGGGTTCGGGATGTTCTATGCCGTGAGCAGGGGCCGCAGGACC  58

Qy   38 GlyValPheLeuThrTrpAsnGluCysArgAlaGlnAlaAspArgPheProAlaAlaArg  57
         ||| ||  ||  ||  | |||||||||  ||||| |||||||  ||||||   ||||||
Db   59 GGGGTCTTTCTGACCTGGAATGAGTGCAGAGCACAGGTGGACCGTGTTTCCTGCTGCCAGA  118

Qy   58 PheLysLysProPheAlaThrGluAlaAspGluAlaTrpAlaPheValArgLysSerAlaSerPro  77
         |||||||| |  |||||| ||   || |  ||  |  |  | |||  |||  |||  || |||
Db  119 TTTAAGAAGTTCCCAGCAGGATGAGGCCTGGCCTTTGTCAGGAAATCTGCAAGCCCG  178

Qy   78 GluValSerGluGlyHisGluAsnGlnHisGluSerGlyLeuGlnAlaGlyProGlyLys  97
         |||||  ||||||||||  ||  ||||  |||||  ||| |||||   ||||| |  |||
Db  179 GAAGTTTCAGAAGGCATGAAAATCAACATGGAAGATCGGACAAGAATCGGAGGCGAGGCAGCCAGCAAG  238

Qy   98 ArgLeuArgGluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaAlaLysHis  117
```

US-10-054-313-1 (1-286) x BQ921568 (1-935)

```
BASE COUNT    235 a    211 c    301 g    182 t
```

Alignment Scores:
```
Pred. No.:              3.77e-115        Length:          935
Score:                  1281.00          Matches:         255
Percent Similarity:     93.17%           Conservative:    4
Best Local Similarity:  91.73%           Mismatches:      11
Query Match:            82.86%           Indels:          8
DB:                                      Gaps:            0
```

```
Qy    1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaAlaLeuProCysArgArg  20
         |||                           |||||  |   |  ||||||  ||  || ||
Db   98 ATGAGCTGGCTTCTTGTCCTGGCCCACAGAGTCGCCTTGGCCGCCTTGCCCTGCCGCCGC  157

Qy   21 GlySerArgGlyPheGlyMetPheTyrAlaAlaValArgArgGlyArgLysThrGlyValPhe  40
         ||| ||| || || |||||||| ||  ||   |   ||||| ||| ||| ||| |||||  ||
Db  158 GGCTCTCGCGGGTTCGGGATGTTCTATGCCGTGAGGAGGGGCCGCAAGACCGGGGTCTTT  217

Qy   41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys  60
         ||  ||||||||||||||  |||||  ||||| |||||||| |  ||||||   ||||||
Db  218 CTGACCTGGAATGAGTGCAGAGCACAGGTGGACCGGTTTCCTGCTGCCAGATTTAAGAAG  277

Qy   61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer  80
         |||||||| |  | | |  |  |   |  |  | |||  || |||  ||||| | |||  |
Db  278 TTTGCCACAGAGGATGAGGCCTGGCCTTTGTCAGGAAATCTGCAAGCCCGGAAGTTTCA  337

Qy   81 GluGlyHisGluAsnGlnHisGluSerGlyLeuGlnAlaGlyProGlyLysArgLeuArg  100
         |||||||||||||||  || |||||  |   | ||| |  ||||| |||||| |||||||
Db  338 GAAGGGCATGAAAATCAACATGGACGAAGAAGCGAGGCGAAAGCCAGCAAGCGACTCCGT  397

Qy  101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro  120
         |||||| |||||| ||||| |  |   | |||  |||||| |   | |||||  || |||
Db  398 GAGCCACTGGATGGAGATGGACATGAAAGCGCAGAGCCGTATGCAAAGCACATGAAGCCG  457

Qy  121 SerValGluProAlaProProValSerArgArgAspThrPheSerTyrMetGlyValAspProPheVal  140
         || |||||||| || ||| |  |  ||  | |  | ||    ||  |  || ||    |  |  |||||
Db  458 AGCGTGGAGCCGGCGCCCTCCAGTAGCAGAGACACGTTTTCCTACATGGGAGACTTCGTC  517

Qy  141 ValValTyrThrAspGlyGlyCysCysSerSerAsnGlyArgArgGlyAlaGlyIle  160
         ||||| || |||||||| ||  ||  |  |   |    |  ||  ||    | |  |||
Db  518 GTCGTCTACACTGATGGCTGCTGCTCCAGTAATGGCCGCCGCGGACCGCGGCGCCGCAGGAATC  577

Qy  161 GlyValTyrTrpTyrGlyProGlyHisProGlyLeuAsnValGlyIleArgGlyLeuProGlyArgGln  180
         ||| ||  | || || ||||| || ||||| |  |  || |  |  |||  ||  ||||||||||  |||
Db  578 GGGGTTTACTGGTGGGGGCCGGGCCATCCTTTANATGTAGGCGTTAAATGGCTTAGAGTTACTAGACTA  637

Qy  181 ThrAsnGlnArgAlaGluIleHisAlaAlaCysGlyAlaAlaIleGluGlnAlaAlaLysThrGln  200
         |||||  |||||||||||  |||||||||  |||  ||||| ||  | | | |||  || |||
Db  638 ACAAACCAAAGAGCGGAAATTCATGCAGCCTGCGAAAGCCATTGAACAAGCAAAGACTCAA  697

Qy  201 AsnIleAsn-LysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAs  220
         ||  |  ||     |  |  |  | |  |  |  ||  |  ||  ||   ||  ||
Db  698 AACATCAATAAAACTGGTTCTGTATCAGACAGTAGTGTTACGATAAATGGTATAACTAA  757

Qy  220 nTrpValGlnGlyTrpProLys-LysAsnGlyTrpLysTrpThrSerAla--GlyLysGluValIle  239
                  ||   | |||          |||||  |  |||  |    |  |
Db  758 CTGGGNGTTCAGGTTGGAAGAAANAAATGGGTGGGAAGACAGTGCANGGGAAAGAGGTGAT  817

Qy  239 eAsnLysGluAspPheValAlaLeu-GluArgLeuThrGlnGlyMet-AspIleGlnTrp  258
         |                ||        || ||  |  |||| |   || |  |  |  |  ||
Db  818 CAACAAGAGGACTTTGTGCACTGGGAGAGCTTAACCCAGGGAATGGGACATCCAGNGGN  877

Qy  259 MetHis-ValProGlyHis-SerGlyPheIleGlyAsn  270
         ||| |  || |||||  |    |   |||  |  | |   | |::
Db  878 ATGCATGGTTCCTGGGCATTCGGGATTAATAGGCCAT  915
```

RESULT 13
BU595053
BU595053
LOCUS   BU595053   818 bp   mRNA   linear   EST 20-SEP-2002
DEFINITION AGENCOURT_8844002 NIH_MGC_142 Homo sapiens cDNA clone IMAGE:6450776
            5' mRNA sequence.
ACCESSION   BU595053
VERSION   BU595053.1 GI:23246812
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 818)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: NCI
            cDNA Library Preparation: Michael Brownstein Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2621 row: p column: 09
            High quality sequence stop: 533.
FEATURES   Location/Qualifiers
     source   1..818
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6450776"
                     /tissue_type="mixed (pool of 40 RNAs)"
                     /lab_host="DH10B (T1-phage-resistant)"
                     /clone_lib="NIH_MGC_142"
                     /note="Vector: pDNR-LIB; Site_1: SfiI (ggccattatggcc);
                     Site_2: SfiI (ggccgcctcggcc);Double-stranded cDNA was
                     prepared from a pool of 40 cell line polyA+ RNAs (bladder
                     - 2%, blood - 33.4%, brain - 5.6%, breast - 12.5%, colon -
                     4%, connective tissue - 1.4%, eye - 1%, intestine - 2.6%,
                     kidney - 2.2%, liver - 5.7%, lung - 10.8%, NK-cell - 5.2%,
                     ovary - 4%, pharynx - 2.5%, prostate - 4.3%, salivary
                     gland - 1.3%, and skin - 2.3%). 5' and 3' adaptors were
                     used in cloning as follows:
                     5'-AAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGG-3' and
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)NN-3'. Full-length
                     enriched library was constructed using the Clontech
                     Creator SMART kit and size-selected to contain the >0.5 kb
                     size fraction (other fractions present in NIH_MGC_141.
                     Library created in the laboratory of M. Brownstein (NIMH,
                     NIH). Note: this is a NIH_MGC Library."

BASE COUNT   204 a   192 c   264 g   155 t   3 others
ORIGIN

Alignment Scores:
Pred. No.:
Score: 1229.00
Percent Similarity: 96.37%
Best Local Similarity: 95.16%
Query Match: 79.50%
DB:

Length: 818
Matches: 236
Conservative: 3
Mismatches: 5
Indels: 4
Gaps: 0

```
US-10-054-313-1 (1-286) x BU595053 (1-818)

Qy    1 MetSerTrpLeuLeuPheLeuAlaLeuHisArgValAlaAlaAlaLeuProCysArgArg   20
Db   67 ATGAGCTGGCTTCTGTTCCTGGCCCTGCACAGAGTCGCCGCTGCCCTGCCGGCGC       126

Qy   21 GlySerArgGlyPheGlyMetPheTyrArgValAlaValArgArgLysThrGlyValPhe   40
Db  127 GGCTCTCGCGGGTTCGGGATGTTCTATCGCGTGGCCGTGAGGAGGAAGACCGGGGTCTTT  186

Qy   41 LeuThrTrpAsnGluCysArgAlaGlnValAsnAspArgPheProAlaAlaArgPheLysLys 60
Db  187 CTGACCTGGAATGAGTGCAGAGCACAGGTGGACCGGTTTCCTGCTGCCAGATTAAGAAG   246

Qy   61 PheAlaThrGluAspGlyAlaTrpAlaPheValArgLysSerAlaSerProGluValSer   80
Db  247 TTTGCCACAGAGGATGAGGGCCTGGGCCTTTGTCAGGAAAATCTGCAAGCCCGAAGTTTCA 306

Qy   81 GluGlyHisGluAsnGlnHisGlyGlnHisGlyLysProGlyLysAlaLysProGlyLeuArg 100
Db  307 GAAGGGCATGAAAATCAACATGCACAAGATCGGAGGCCAAGGCCAAGCGACTCCGT     366

Qy  101 GluProLeuAspGlyHisGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysLysPro 120
Db  367 GAGCCACTGGATGGAGATGGACATGAAAGCGCACAGCCGTATGCAAAGCACATGAAGAAGCCG 426

Qy  118 MetLysProSerValGluProAlaProProValSerArgAspArgPheSerTyrMetGly  137
Db  299 ATGAAGCCAAGCGTGGAGCCGGCGCCTCCAGTAGCAGAGACACGTTTTCCTACATGGGA   358

Qy  121 SerValGluProAlaProProValSerArgAspArgPheSerTyrMetGlyAspPheVal  140
Db  427 AGCGTGGAGCCGGCGCCTCCAGTAGCAGAGACACGTTTTCCTACATGGGAGACTTCGTC   486

Qy  138 AspPheValValValTyrThrArgGlyCysSerSerSerAsnGlyArgArgLysProArg  157
Db  359 GACTTCGTCGTCGTCTACACTAGGTGCTGCTCCTCCAGTAATGGGCGTAGAAGGCCGCGA  418

Qy  141 ValValTyrThrArgGlyCysSerSerSerAsnGlyArgArgLysProArgAlaGlyIle  160
Db  487 GTCGTCTACACTAGGTGCTGCTCCTCCAGTAATGGGCGTAGAAGGCCGCGAGCAGGAATC  546

Qy  158 AlaGlyIleGlyValTyrThrGlyTrpProGlyHisProLeuAsnValGlyIleArgLeuPro 177
Db  419 GCAGGAATCGGCGTTTACTGGGGCCAGGCCATCCTTTAAATGTAGGCATTAGGACTTCCT  478

Qy  161 GlyValValTyrTrpTrpGlyValHisProGlyHisProLeuAsnValGlyIleGlyLeuArgProGlyArgGln 180
Db  547 GGCGTTTACTGGTGGGGCGTGCACCCCGGCCATCCTTAAATGTAGGCATTAGGACTTCCTCGGGGCGGCCAG  606

Qy  178 GlyArgGlnThrAsnGlnArgAlaAlaGluIleHisAlaAlaCysLysAlaIleGluGlnAla 197
Db  479 GGGCGGCCAGACAAACCAAAGAGCGGAAATTCATGCAGCCTGCAAAGCCATTGAACAACAGCA 538

Qy  181 ThrAsnGlnArgAlaAlaGluIleHisAlaAlaCysLysAlaIleAlaIleGluGlnAlaLysThrGln 200
Db  607 ACAAACCAAAGAGCGGAAATTCATGCAGCCTGCAAAGCCATTGAACAGCAGCCAAAGACCTCAA  666

Qy  198 LysThrGlnAsnIleAsnLysAsnValLeuTyrThrAspSerMetPheThrIleAsnGly   217
Db  539 AAGACTCAAAACATCAATAAACTGGTTCTGTATACAGACAGTATGTTTACGATAAATGGT   598

Qy  201 AsnIleAsnLysAsnValLeuValTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220
Db  667 AACATCAATAAAAACTGGGTCTGTATACAGACAGTATGTTTACGATAAATGGTATAACTAAC  726

Qy  218 IleThrAsn-TrpValGlnGlyTrpProLysLysAsnGlyTyrTrpLysThrSerAlaGlyLysGly 237
Db  599 ATAACTAACGTGGGTTGGAAGACAAATGGGTGGAAGACAAGTGCAGGGAAAGA       658

Qy  221 TrpVal-GlnGlyTrpProLysLysAsnGlyTrp-LysThrSerAla-GlyLysGluValIleAsnLysGluValIleIle 239
Qy  237 uValIleAsnLysGluAspPheValAlaLeuGlu-ArgLeuThrArgGlnGlyMetAspIleG 257
Db  659 GGTGATCACAAAGAGGACTTTGTTGCACTGGAGAAGGCTTACCCAGGGATGGACATTC    718

Qy  257 lnTrpMetHisValProGlyHisSerGlyPhe 267
Db  719 AGTGGATGCT-GTTCCTGGTCATTCGGGATTT 749
```

```
Db    727  TGGGTTCCAGGGTTGGAAGAANAAATGGGTGGNAAGACNAGTGCAGGGGAAAGAGGTGATC  786

Qy    240  Asn-LysGluAspPhe  244
                :::  |||
Db    787  CACAAAAGAGGACTTT  802

RESULT 14
BM451915
LOCUS       BM451915          916 bp    mRNA        linear    EST 05-FEB-2002
DEFINITION  AGENCOURT_6397386 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:5492781
            5', mRNA sequence.
ACCESSION   BM451915
VERSION     BM451915.1  GI:18500955
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 916)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
  COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12115 row: g  column: 22
            High quality sequence stop: 626.
FEATURES             Location/Qualifiers
     source          1..916
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5492781"
                     /tissue_type="retinoblastoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_67"
                     /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.75 kb.  Library constructed by Life
                     Technologies."

BASE COUNT
ORIGIN      223  a      229  c      282  g      180  t      2 others

Alignment Scores:
Pred. No.:
Score:                    2.2e-109              Length:          916
Percent Similarity:       1222.00               Matches:         247
Best Local Similarity:    89.08%                Mismatches:      6
Query Match:              86.97%                Conservative:    16
DB:                       79.04%                Indels:          15
                          12                    Gaps:            4

US-10-054-313-1 (1-286) x BM451915 (1-916)

Qy      1  MetSerTrpLeuLeuPheLeuAlaHisArgValAlaAlaLeuAlaAlaLeuProCysArgArg  20
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     75  ATGAGCTGGCTGCTTTTCCTGGCCCACAGAGTCGCCTTGGCCGCCTTGGCCTGCCGCGGCGC  134

Qy     21  GlySerArgGlyPheGlyMetPheTyrArgValArgArgGlyArgGlyLysThrGlyValPhe  40
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    135  GGCTCTCGCCGCGGTTCGGATGTTCTATGCCGTGAGGAGGGGCCCGAAGAACCGGGGTCTTT  194

Qy     41  LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysGlyLys  60
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    195  CTGACCTGGAATGAGTGCAGAGACAGGTGGACCCGGTTTCTGTCTGCCAGATTTAAGAAG  254

Qy     61  PheAlaThrGluAlaGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer  80
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    255  TTTGCCACAGAGGATGAGGCCTGGGCCTTTGTCAGGAAATCTGCAAGCCCGGAAGTTTCA  314

Qy     81  GluGlyHisGluAsnGlnHisGlyLyGlnGluSerAlaLysProGlyLyArgLeuArg  100
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    315  GAAGGGCATGAAAATCAACATGGACAAGAATCGGAGCGAGCAAGCCAGCAAGCGACTCCGT  374

Qy    101  GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro  120
            |||||||||||||||||||||||||||||||||:::||||||||||||||||||||||
Db    375  GAGCCACTGGATGGAGATGGACATGAAAGGCCAGAGCCGTATGCAAAGCACATGAAGCCG  434

Qy    121  SerValGluProAlaProValSerArgArgAspThrPheSerTyrMetGlyAspPheVal  140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    435  AGCGTGGAGCCGGCGGCCCTCCAGTTAGCAGAGACACGTTTCCTACATGGAGACTTCGTC  494

Qy    141  ValValTyrThrAspGlyCysSerSerArgAsnGlyArgArgLysProArgAlaGlyIle  160
            |||||||||||||||||||||||||||||:::||||||||||||||||||||||||||
Db    495  GTCGTCTACACTGATGGCTGCTCCAGTAATGGCGTAGAAGGGCCGGAGCAGGAATC  554

Qy    161  GlyValTyrTrpGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGln  180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    555  GGGGTTTACTGGGGGCCAGGCCATCCTTTAAATGTAGGCATTAGACTTCCTGGCCGCAG  614

Qy    181  ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaAlaIleGluGlnAlaLysThrGln  200
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    615  ACAAACCANAGAGCCGAAATTCATGCAGCCTGCAAAGCCATTGAACAAGCAAAGACTCAN  674

Qy    201  AsnIleAsn-LysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAs  220
            ||||:::|||||||||||||||||||||||||||||||||||||||||||||||||||
Db    675  AACATCCATAAAACTGTTCTGTTATACAGACAGTAGTTTACGATAAATGGTATAACTAA  734

Qy    220  nTrp-ValGlnGly-TrpLysLysAsn-GlyTrpLysThrSerAla-GlyLysGluVal  239
            |||:::|||||||::||||||||||:||||||||||||||||||:||||||||||
Db    735  CTGGGGTTCCAGGTTTGGAAGAAAAATGGGTGGAAGACCAGTGCAGGGGAAAAGAGGTGA  794

Qy    239  leAsn--LysGluAspPheValAlaLeuGluLeuThrGlnGlyMetAspIleGlnT  258
            |||:::            |||::|||                    |||
Db    795  TTCACCAAGAGGACTTTGTGGCACTGGAAAGGCTCTACCCAGGG--------GGAT  845

Qy    258  rpMetHis--------------ValProGlyHisSerGly---PheIleGlyAsnGluG  272
            |||                   |||||||||||:||:::
Db    846  GGACCTTTCCCCGGGATGCTTGTTCCCGGGGCCATCCGGGGCCATTTTTAGGCCAATGAAA  905

Qy    272  lu 272
            ::
Db    906  AA 907
```

```
RESULT 15
BF984400
LOCUS       BF984400          869 bp    mRNA        linear    EST 23-JAN-2001
DEFINITION  602307923F1 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:4399301 5',
            mRNA sequence.
ACCESSION   BF984400
VERSION     BF984400.1  GI:12387212
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 869)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
  COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10102 row: n  column: 06
            High quality sequence stop: 691.
```

```
FEATURES          Location/Qualifiers
     source       1..869
                  /organism="Homo sapiens"
                  /mol_type="mRNA"
                  /db_xref="taxon:9606"
                  /clone="IMAGE:4399301"
                  /tissue_type="duodenal adenocarcinoma, cell line"
                  /lab_host="DH10B (phage-resistant)"
                  /clone_lib="NIH_MGC_88"
                  /note="Organ: small intestine; Vector: pCMV-SPORT6;
                  Site 1: NotI; Site 2: SalI; Cloned unidirectionally;
                  oligo-dT primed. Average insert size 1.767 kb. Library
                  enriched for full-length clones and constructed by Life
                  Technologies. Note: this is a NIH_MGC Library."

BASE COUNT    249 a    177 c    270 g    173 t
ORIGIN
```

```
Alignment Scores:
Pred. No.:                          Length:         869
Score:            1215.00           Matches:        231
Percent Similarity: 97.90%          Conservative:     2
Best Local Similarity: 97.06%       Mismatches:       4
Query Match:       78.59%           Indels:           2
DB:                                 Gaps:             0
```

```
US-10-054-313-1 (1-286) x BF984400 (1-869)

Qy    47 ArgAlaGlnValAspArgPheProAlaAlaArgPheLysPheLysPheAlaThrGluAspGlu   66
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3 AGAGCACAGGTGACCGGTTTCCTGCTGCCAGATTTAAGAAGTTTGCCACAGAGGATGAG       62

Qy    67 AlaTrpAlaPheValArgLysSerAlaSerProGluValSerProGluGlyHisGluAsnGln   86
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    63 GCCTGGGCCTTTGTCAGGAAATCGCAAGCCCGGAAGTTTCAGAAGGGCATGAAAATCAA     122

Qy    87 HisGlyGlnGluSerGlyLysAlaLysProGlyLysArgLeuArgGluLeuAspGlyAspAsp  106
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   123 CATGGACAAGAATCGGAGGCGAAAGCCAGCAGCGCGACTCCGTGAGCCACTGGATGGAGAT   182

Qy   107 GlyHisGluSerAlaGlnProTyrAlaLysHisMetLysProSerValGluProAlaPro    126
          ||||||||||| |||:::||||||||||||||||||||||||||||||||||||||||||
Db   183 GGACATGAAAGCCCAGAGCCGTATGCAAAGCACCACATGAAGCCCAGCGTGGAGCCGGCCCT   242

Qy   127 ProValSerArgAspThrPheSerTyrPheSerTyrMetGlyAspPheValValTyrThrAspGly 146
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   243 CCAGTTAGCAGAGACACGTTTTCCTACATGGGAGACTTCGTCGTCGTCTACACTGATGGC   302

Qy   147 CysCysSerSerAsnGlyArgArgLysProArgAlaGlyTyrIleGlyTyrValTyrTrpGlyPro 166
          |||||||||||||||| :::|||||||||||||||||||||||||||||||||||||||
Db   303 TGCTGCTCCAGTAATGGGCGTAGAAGCCGCAGGAGCGTTTACTGGGGGCCG         362

Qy   167 GlyHisProLeuAsnValGlyTyrIleArgLeuProGlyArgGlnThrAsnGlnArgAlaGlu  186
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   363 GGCCATCCTTTAAATGTAGGCATTAGACTTCCTGGGCGGCAGACAAACCAAAGAGCCGAA     422

Qy   187 IleHisAlaAlaCysLysAlaIleGluGlnAlaLysThrGlnAsnIleAsnLysLeuVal    206
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   423 ATTCATGCAGCCTGCAAAGCCATTGAACAAGCAAAGACTCAAAACATCAATAAACTGGTT   482

Qy   207 LeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsnGlyValGlnGlyTrpTrpLys  226
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   483 CTGTATACAGACAGTGTTCAGATAAATGGTATAACTAACTGGGTTCAAGGTTGGAAG       542

Qy   227 LysAsnGlyTrpLysTyrSerArgAlaGlyLysGluIleLeuAsnGlyAspPheValAla    246
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   543 AAAAATGGGTGGAAGACAAGTGCAGGGAAAGAGTGATCAACAAAGAGGACTTTGTGGCA     602

Qy   247 LeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHisValProGlyHisSerGly   266
          ||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db   603 CTGGAGAGGCTTACCCA-GGGATGGACATTCAGTGGATGCATGTTCCTGGTCATTCGGGA   661

Qy   267 PheIleGlyAsnGluGluAlaAspArgLeuAlaArgGluGlyAlaLysGln      283
          ||||||| |||||||||||||  |||||||||||
```

```
Db   662 TTTATAGGCAATGAAGAAGCTGACAGATTAGGCAGAGAGAGAGGAGGCTAAACAA  713

Search completed: December 17, 2003, 21:11:15
Job time : 2528 secs
```